

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université des Sciences et de la Technologie Houari Boumediene
Faculté d'Electronique et d'Informatique



THESE

Présentée pour l'obtention du grade de Doctorat En Sciences

En : Electronique

Spécialité : Communication Parlée

Par : SAADOUNE Adda

THÈME

**Rehaussement de la Parole par les Méthodes
PCA-VRE**

Soutenue publiquement le 16 Décembre 2014 devant le Jury composé de:

Mr A. GUESSOUM	Professeur, à l'USD Blida	Président
Mr A. AMROUCHE	Professeur, à l'USTHB	Directeur de Thèse
Mr S.A. SELOUANI	Professeur, à l'UMCS Canada	Co-Directeur de Thèse
Mr A. HOUACINE	Professeur à l'USTHB	Examineur
M^{me} M. GUERTI	Professeur, à l'ENP Alger	Examinatrice
Mr H. SAYOUD	Professeur à l'USTHB	Examineur

Résumé

Mots Clefs : KLT, PCA, PKLT, VRE, MCRA, SVD et EVD.

Le rehaussement de la parole qui consiste à améliorer certains aspects perceptifs du signal vocal dégradé par un bruit, le plus souvent de nature additif, s'est élargi au cours de ces dernières années sur plusieurs domaines d'application du traitement du signal, notamment ceux en rapport avec la téléphonie mobile et IP, l'analyse des signaux biomédicaux, les télécommunications militaires, la restauration d'enregistrements audio anciens, etc.

Par ailleurs, le développement d'une méthode de rehaussement nécessite a priori des connaissances sur le comportement des différents types de bruits qui, dans beaucoup de cas, sont des processus non stationnaires.

Le rehaussement de la parole peut être abordé par plusieurs méthodes du traitement du signal et de l'algèbre linéaire, à savoir : la soustraction spectrale, le filtrage de Wiener, le filtrage de Kalman, les méthodes basées sur des modèles statistiques, les méthodes basées sur la décomposition en sous-espaces ainsi que celles fondées sur des modèles perceptuels.

Idéalement, les algorithmes de rehaussement doivent être en mesure d'améliorer à la fois la qualité et l'intelligibilité du signal. Cependant, en pratique, la plupart de ces algorithmes agissent uniquement sur la qualité, par réduction du bruit additif, tout en introduisant des distorsions sur le signal rehaussé qui affectent son intelligibilité.

Nous avons proposé dans le cadre de cette thèse, deux méthodes de rehaussement de la parole basées sur la décomposition en sous-espaces, dénommées "KLT-VRE-MCRA" et "PKLT-VRE" que nous avons testées sur un corpus de phrases phonétiquement équilibrées bruitées à des niveaux de Rapport Signal sur Bruit de 0, 5 et 10 dB, issues de la base de données NOIZEUS.

La première méthode proposée consiste à intégrer avec la méthode "KLT-VRE", considérée dans notre cas comme une méthode de référence, une technique d'estimation de la Densité Spectrale du Bruit (DSP) appelée MCRA (Minima Controlled Recursive Averaging) qui consiste à traquer la valeur minimale de la puissance du signal bruité dans chaque trame d'analyse. La combinaison de la Variance de l'Erreur de Reconstruction (VRE : Variance of the Reconstruction Error) et de la technique MCRA avec le modèle KLT (Karhunen-Loève

Transform) vise d'une part, à optimiser le modèle KLT par la sélection du nombre optimal de Composantes Principales (PC) devant être utilisées dans la reconstruction du signal propre et d'autre part, à prendre en charge effectivement les bruits non stationnaires en vue de se rapprocher au mieux du cas réel. Nous avons utilisé trois variantes de MCRA, en l'occurrence MCRA, MCRA2 et IMCRA (Improved MCRA), à l'effet de confirmer la robustesse de la méthode proposée

La deuxième méthode proposée quant à elle, a été développée à partir de la mise en parallèle des méthodes PKLT et KLT-VRE. La nouvelle fonction de suppression proposée est une combinaison des deux matrices de gains respectifs qui a été optimisée ensuite par une fonction de régression linéaire. Les résultats obtenus témoignent de la robustesse de la méthode proposée du fait qu'elle permet de tirer profit des propriétés perceptuelles du système auditif en vue d'améliorer à la fois la qualité et l'intelligibilité de la parole rehaussée.

Plusieurs tests comparatifs ont été conduits avec des méthodes de l'état de l'art, à l'effet d'évaluer objectivement les performances des deux méthodes proposées, et ce, à travers une technique numérique qui se substitue à la technique classique du jugement et qui permet d'évaluer le signal de la parole propre (SIG : Signal Distortion), le bruit (BACK : Background Noise) ainsi que la qualité globale de la parole (OVL : Overall Quality). Les mesures objectives, utilisées dans le cadre de cette thèse, ont concerné également l'évaluation du niveau de distorsion entre le signal propre et le signal rehaussé, par le biais des techniques PESQ (Perceptual Evaluation of Speech Quality), WSS (Weighted-Slope Spectral), SegSNR (Segmental SNR) et LLR (Log-Likelihood Ratio).

Les résultats obtenus, à l'issue de ces tests, ont été dans leurs majorités probants, ce qui confirme la robustesse des méthodes proposées.

Remerciements

Je tiens tout d'abord à exprimer mes remerciements les plus sincères à Monsieur Abderrahmane AMROUCHE, mon directeur de thèse, pour son encouragement, son accompagnement, son soutien et sa disponibilité tout au long de la thèse ainsi que pour la confiance qu'il m'a accordé et ses judicieux conseils qui ont contribué à alimenter ma réflexion.

J'adresse mes sincères remerciements à mon co-directeur de thèse, Monsieur Sid Ahmed SELOUANI de l'UMCS / Canada pour avoir accepté de m'accueillir au sein de son Laboratoire de Recherche en Interaction Humain Système "LARIHS" et que sans son concours effectif, ce travail n'aurait jamais vu le jour. Je tiens à le remercier encore une fois pour tout ce qui a fait pour moi durant mon séjour à l'UMCS.

Je remercie également Madame et Messieurs les professeurs : M. GUERTI de l'ENP, A. GUESSOUM de l'USD/Blida, A. HOUACINE de l'USTHB, A. BELCOUHRANI de l'ENP et H. SAYOUD de l'USTHB de m'avoir fait l'honneur de participer à ce jury de thèse.

Mes remerciements vont également à tous mes amis et mes proches qui m'ont soutenu durant ces années d'études.

Pour terminer, je voudrais aussi ajouter mes remerciements à toute ma famille pour m'avoir toujours encouragé et soutenu moralement tout au long de mon parcours universitaire.

Sommaire

Résumé	i
Remerciements	iii
Sommaire	iv
Liste des abréviations	vii
Liste des figures	ix
Liste des tableaux	xi
Chapitre I : Introduction générale	1
I.1. Introduction	1
I.2. Motivation	3
I.3. Objectifs de la thèse.....	4
I.4. Organisation de la thèse.....	5
Chapitre II : Techniques de rehaussement de la parole	6
II.1. Introduction.....	6
II.2. Méthode de Soustraction Spectrale (SS).....	6
II.3. Filtrage de Wiener Paramétrique.....	8
II.4. Phénomène du bruit musical.....	9
II.5. Techniques de Soustraction Spectrale (SS).....	12
II.5.1. Technique de Soustraction Spectrale Non Linéaire : NSS-SB.....	13
II.5.2. Technique de Soustraction Spectrale Multi-Bandes : MB-SB.....	13
II.5.3. Technique de Soustraction Spectrale basée sur le critère MMSE.....	15
II.5.4. Technique de Soustraction Spectrale basée sur le Moyennage Adaptative du Gain (AGA).....	15
II.6. Méthodes de rehaussement de la parole basées sur des modèles statistiques.....	17
II.6.1. Estimateurs du Maximum de Vraisemblance (ML)	17
II.6.2. Estimateurs Bayésien.....	21
II.6.3. Estimateur de l'Erreur Quadratique Moyenne (MSE).....	22
II.6.4. Estimateur LOG-MMSE (LSA).....	28
II.6.5. Estimateur du Maximum à Posteriori (MAP).....	29
II.6.6. Estimateur MMSE d'ordre p	30
II.6.7. Estimateurs MMSE basés sur des distributions non gaussiennes.....	30
II.6.8. Estimateurs Bayésien Perceptuels.....	30

II.7. Méthodes de rehaussement incorporant la probabilité d'absence de la parole.....	30
II.8. Rehaussement de la parole par filtrage de Kalman.....	33
II.9. Conclusion.....	36
Chapitre III : Rehaussement de la parole par décomposition en sous-espaces du signal	37
III.1. Introduction.....	37
III.2. Formalismes mathématiques.....	38
III.2.1. Définitions.....	38
III.2.2. Matrice de covariance.....	39
III.2.3. Valeurs propres et vecteurs propres.....	40
III.3. Rehaussement de la parole par Décomposition en Valeurs Propres (EVD).....	40
III.4. Estimateur linéaire du signal.....	42
III.4.1. Estimateur TDC.....	43
III.4.2. Estimateur SDC.....	45
III.5. Rehaussement de la parole par Décomposition en Valeurs Singulières (SVD).....	47
III.6. Rehaussement de la parole par La Transformée de Karhunen-Loève (KLT).....	52
III.7. Rehaussement de la parole par Analyse en Composantes Principales (PCA).....	55
III.8. Rehaussement de la parole par KLT-VRE.....	57
III.8.1. Critères AIC et MDL.....	57
III.8.2. Critère IEF.....	59
III.8.3. Critère VRE.....	59
III.9. Conclusion.....	64
IV. Méthodes PKLT-VRE et KLT-VRE-MCRA de rehaussement de la parole proposées	65
IV.1. Introduction.....	65
IV.2. Définition des paramètres communs.....	66
IV.2.1. Corpus de tests.....	66
IV.2.2. Evaluation des performances.....	66
IV.3. Méthode de rehaussement de la parole "KLT-VRE-MCRA" proposée.....	67
IV.3.1. Description de la méthode MCRA.....	67
IV.3.2. Description de la méthode "KLT-VRE-MCRA" proposée.....	69
IV.3.3. Résultats et discussions.....	70
IV.4. Méthode de rehaussement de la parole "PKLT-VRE" proposée.....	73
IV.4.1. Description de la méthode "PKLT-VRE" proposée.....	76

IV.4.2. Optimisation de la matrice du gain.....	76
IV.4.3. Résultats et discussions	77
IV.5. Conclusion.....	86
Conclusions et perspectives	87
1. Conclusions.....	87
2. Perspectives.....	89
Références bibliographiques	90

Liste des abréviations

AGA	Adaptive Gain Averaging
AIC	Akaike Information Criterion
BAK	Background Intrusiveness
DSP	Densité Spectrale de Puissance
EVD	Eigen Value Decomposition
FET	Frequency Eigen-domain Transformation
FDP	Fonction Densité de Probabilité
IEF	Imbedded Error Function
IKLT	Inverse KLT
IMCRA	Improved MCRA
KLT	Karhunen-Loève Transform
LS	Least Square
LLR	Log Likelihood Ratio
LSA	Log Spectral Amplitude
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MCRA	Minima Controlled Recursive Averaging
MB	MultiBand Spectral Subtraction
MAP	Maximum a Posteriori
MDL	Minimum Description Length
MOS	Mean Opinion Score
NSS	Non linear Spectral Subtraction
OVL	Overall Quality
PKLT	Perceptual Karhunen-Loève Transform
PCA	Principal Component Analysis

PCS	Principal Component Space
PESQ	Perceptual Evaluation of Speech Quality
RS	Residual Space
RSB	Rapport Signal sur Bruit
SVD	Singular Value Decomposition
STSA	Short-Time Spectral Amplitude
SDC	Spectral Domain Constrained
SIG	Signal Distortion
SegSNR	Segmental Signal to Noise Ratio
SS	Soustraction Spectrale
TFD	Transformée de Fourier Discrète
TDC	Time Domain Constrained
TF	Transformée de Fourier
TFCT	Transformée de Fourier à Court Terme
VRE	Variance of the Reconstruction Error
WSS	Wide Sens Stationary
WSE	Wide Sens Ergodic
WSS	Weighted Slope Spectral

Listes des figures

Fig. II. 1	Représentation temporelle d'un signal propre, corrompu par un bruit de chahut à 0B de RSB et rehaussé par la méthode de Soustraction Spectrale de Berouti [18].....	11
Fig. II. 2	Courbes d'atténuations de l'algorithme de Soustraction Spectrale pour différentes valeurs de α [1].....	12
Fig. II. 3	Schéma bloc de l'algorithme de Soustraction Spectrale MB.....	14
Fig. II. 4	Spectrogrammes d'une phrase issue de la base de donnée NOIZEUS avec (a) signal propre, (b) signal bruité avec du chahut à 5dB, (c) signal rehaussé par SS_Berouti (d) signal rehaussé par SS_ MB et (e) signal rehaussée par SS_ AGA.....	16
Fig. II. 5	Courbes d'atténuations de la méthode du maximum de vraisemblance [1].....	21
Fig. II. 6	Tracer de la probabilité à posteriori de présence de la parole $p(H_1^k X_k)$, en fonction du RSB à posteriori Y_k [1].....	32
Fig. III.1	Estimateur linéaire de la méthode de décomposition en sous-espace [1].....	45
Fig. III.2	Représentation des valeurs singulières dans le sous-espace du signal et sous-espace du bruit [52].....	52
Fig. IV.1	Schéma bloc de la méthode KLT-VRE-MCRA proposée [50].....	69
Fig. IV.2	Spectrogramme d'une phrase de la base de données NOIZEUS, corrompue par un bruit de chahut et rehaussée par différentes méthodes de rehaussement comme suit : (a) parole propre (b) parole corrompue par le bruit de chahut, (c) méthode KLT-VRE-MCRA proposée, (d) méthode KLT-VRE, (e) méthode LSA, (f) méthode de Wiener et (g) méthode SS.....	72
Fig. IV.3	Schéma bloc de la méthode PKLT-VRE proposée [51].....	77
Fig. IV.4	Spectrogrammes des méthodes de rehaussement de la parole : (a) parole propre, (b) parole bruité dégradée par un bruit chahut à 0 dB de RSB d'entrée, (a1 à d1) Spectrogrammes de la méthode de SS pour différentes valeurs de F et γ ; (a2 à d2) Spectrogrammes de la méthode de Wiener pour différentes valeurs de α et β ; (a3 à d3) Spectrogrammes de la méthode KLT-VRE pour différentes valeurs de μ_L ; (a4 à d4) Spectrogrammes de la méthode PKLT pour différentes valeurs de μ_L et ν ; (a5 à d5) Spectrogrammes de la méthode proposée PKLT-VRE proposée pour différentes valeurs de μ_L et ν	80

Fig. IV.5 Evaluations objectives avec PESQ, SegSNR et WSS pour different methods en utilisant : (a) bruit de "bus" avec NOIZEUS, (b) bruit de "rue" avec NOIZEUS, (c) bruit de "chahut" avec TIMIT..... 81

Liste des tableaux

Tab. II.1	Filtrage de Wiener paramétrique et méthodes correspondantes pour quelques valeurs de (δ, ρ)	9
Tab. IV.1	Evaluations objectives (PESQ et RSB) de la méthode KLT-VRE-MCRA comparées avec les méthodes de Wiener, SS, LSA et KLT-VRE.....	71
Tab. IV.2	Evaluation objective de la méthode PKLT-VRE proposée comparée avec SS, Wiener, PKLT et KLT-VRE et corrompue par des bruits de bus, rue et chahut. Valeurs moyenne obtenues en utilisant 20 phrases extraites de la base de données NOIZEUS. Les meilleures performances sont indiquées en Gras.....	83
Tab. IV.3	Evaluation objective en utilisant des données disjointes composées de 10 phrases extraites des bases de données NOIZEUS et TIMIT. La méthode PKLT-VRE proposée est comparée avec SS, Wiener, PKLT et KLT-VRE et corrompue avec le bruit de chahut. Les meilleures performances sont indiquées en Gras.....	84
Tab. IV.4	Evaluation objective en utilisant des données disjointes composées de 10 phrases extraites des bases de données NOIZEUS et TIMIT. La méthode PKLT-VRE proposée est comparée avec SS, Wiener, PKLT et KLT-VRE et corrompue avec le bruit d'avion de chasse. Les meilleures performances sont indiquées en Gras.....	84
Tab. IV.5	Evaluation objective en utilisant des données disjointes composées de 10 phrases extraites des bases de données NOIZEUS et TIMIT. La méthode PKLT-VRE proposée est comparée avec SS, Wiener, PKLT et KLT-VRE et corrompue avec le bruit d'usine. Les meilleures performances sont indiquées en Gras.....	85
Tab. IV.6	Evaluation objective de la méthode PKLT-VRE proposée avec optimisation du gain. Cette comparaison est utilisée avec 3 phrases issues de la base de données TIMIT.....	85

Tab. IV.7 les valeurs-P ($p < 0.05$) des mesures objectives SIG, PESQ et SegSNR de la technique PKLT-VRE comparée à SS, Wiener, PKLT et KLT_VRE pour différents types de bruits à 0 dB de RSB d'entré..... 86

Chapitre I

Introduction générale

I.1. Introduction

Le rehaussement de la parole est un domaine de recherche très actif qui s'est largement développé durant ces trois dernières décennies. Sa nécessité s'est manifesté en particulier dans de nombreuses situations, dans lesquelles le signal vocal provient soit d'une source bruitée, soit il est affecté directement par le bruit du canal des transmissions.

En effet, le rehaussement peut être utilisé comme un prétraitement au niveau des codeurs de la parole, mis en place dans les systèmes de radiocommunications de types GSM, 3G, 4G, etc. Si ces systèmes, soumis à des bruits environnementaux typiques (véhicule, restaurant, car, rue, etc.), sont équipés de dispositifs de Reconnaissance Automatique de la Parole (RAP) (exemple : numérotation vocale), le taux de reconnaissance se voit donc décroître assez rapidement en présence de ces types de bruits [1]. Pour remédier à ces contraintes, plusieurs techniques de RAP ont été développées sur la base d'hypothèses de présence du bruit [2-5] et ce, en vue d'améliorer davantage le taux de reconnaissance global.

Aussi, dans le cas des communications vocales sol-air, le rehaussement de la parole permet d'améliorer la qualité et l'intelligibilité de la parole du pilote qui est mélangée avec le bruit présent dans le cockpit. Pour leur part, les systèmes de téléconférence sont caractérisés par le phénomène de réverbération de la salle qui provoque la diffusion du bruit de la source sur toute la salle. Pour cela, on procède au rehaussement de la parole avant sa transmission.

Enfin, dans le cadre de la réhabilitation auditive des malentendants, le rehaussement de la parole permet de maintenir au plus haut niveau les performances des prothèses auditives qui diminuent assez rapidement dans des milieux fortement bruités.

Idéalement, les algorithmes de rehaussement doivent être en mesure d'améliorer la qualité et l'intelligibilité du signal. Cependant, en pratique, ces algorithmes agissent uniquement sur la qualité par la réduction du bruit additif et tout en introduisant des distorsions qui affectent plus ou moins l'intelligibilité du signal reçu. Pour cela, l'enjeu principal dans la conception d'un algorithme efficace de rehaussement de la parole consiste à supprimer le bruit sans pour autant introduire des distorsions perceptibles dans le signal rehaussé.

Par ailleurs, la mise en œuvre d'une méthode de rehaussement de la parole nécessite à priori des connaissances sur le comportement du bruit [1], [6] (niveau de pression sonore, Rapport Signal à Bruit, masquage du bruit, etc.). Aussi, l'estimation du bruit peut avoir un impact majeur sur la qualité du signal rehaussé, au sens où si cette estimation est trop faible, le bruit résiduel devient audible et si elle est trop élevée, la parole sera distordue engendrant ainsi une perte d'intelligibilité.

La méthode la plus simple consiste à estimer et mettre à jour le spectre du bruit pendant les segments de silences du signal en utilisant un algorithme de détection d'activité vocale (VAD) [7-8]. Même si une telle approche peut fonctionner de manière satisfaisante dans le cas d'un bruit stationnaire, elle demeure cependant inadaptée pour des environnements plus réalistes, où les caractéristiques spectrales du bruit changent constamment. Plusieurs méthodes sont donc proposées pour prendre en charge les bruits non stationnaires, et ce, par le biais d'une estimation continue [9-17].

Les méthodes de rehaussement de la parole peuvent être réparties en trois principales catégories à savoir :

1. Méthodes de soustraction spectrale : Le principe de ces méthodes [18-20] consiste à soustraire la valeur estimée du spectre de bruit à partir de la valeur estimée du spectre du signal bruité. Le spectre de bruit, supposé additif, peut être calculé durant la période d'absence d'activité vocale. Le problème majeur avec cette technique est l'apparition d'un bruit résiduel, appelé bruit musical, qui subsiste dans le signal rehaussé après soustraction spectrale et qui est très gênant du point de vue perceptif. Son spectre à court terme correspond approximativement à une distribution aléatoire de pics spectraux ;

2. Méthodes basées sur des modèles statistiques : Le rehaussement de la parole peut être considéré comme un problème d'estimation statistique où pour un ensemble de mesures correspondant aux coefficients de la transformée de Fourier du signal bruité, nous cherchons à trouver un estimateur linéaire (ou non linéaire) des paramètres du signal propre. Ce domaine qui a été initié par McAulay et Malpass [21], a porté sur la proposition d'une approche du Maximum de Vraisemblance (ML) pour l'estimation des coefficients de la transformée de Fourier du signal propre. Ephraim et Malah [10, 21-29] ont, pour leur part, proposé un estimateur de l'erreur quadratique moyenne (MMSE) du spectre d'amplitude.

3. Méthodes de décomposition en sous-espace : contrairement aux méthodes précédentes, les méthodes de décomposition en sous-espaces ont été développées sur la base de la théorie de l'algèbre linéaire [1], [30-31]. Leur principe réside sur le fait que le signal propre peut être limité à un sous-espace de l'espace euclidien bruité. En tenant compte de la méthode qui consiste à décomposer l'espace vectoriel du signal bruité en un sous-espace occupé par le signal propre et un autre sous-espace occupé par le bruit, on peut estimer le signal propre par l'annulation de la composante du vecteur bruité résidant dans le sous-espace de bruit. Cette décomposition en sous-espaces peut être effectuée par des techniques de factorisation des matrices orthogonales appelées Décomposition en Valeurs Singulières (SVD) ou Décomposition en Valeurs Propres (EVD) [32-39]. Ce domaine a été initié par les travaux de Dendrinos et al. [40], qui ont proposé l'utilisation de SVD sur une matrice de données contenant des valeurs d'amplitude dans le domaine temporel. Ensuite, Ephraim et Van tree [32] ont proposé l'utilisation de la EVD de la matrice de covariance du signal. La méthode de décomposition en sous-espace a été également utilisée sur des modèles perceptuels, et ce, par la prise en compte des propriétés perceptuelles du système auditif humain [41-49] afin de rendre le bruit résiduel perpétuellement inaudible. Dans ce cas, l'estimation spectrale du signal de parole joue un rôle capital dans la détermination de la valeur du seuil de masquage du bruit, utilisé pour adapter le facteur du gain perceptuel. Les résultats obtenus montrent un meilleur compromis entre la minimisation de la distorsion du signal et la suppression du bruit musical ;

I.2. Motivation

Cette thèse propose une contribution pour l'amélioration, au sens de la qualité et de l'intelligibilité de la parole perçue, de deux méthodes de décomposition en sous-espaces.

Nous allons tout d'abord examiner ce domaine de recherche en nous focalisant sur la méthode de rehaussement en sous-espaces que nous avons associé d'une part, avec le critère VRE (Variance of the Reconstruction Error) de sélection du nombre optimal de Composantes Principales (PC) à retenir pour la reconstruction du modèle et d'autre part, avec la méthode MCRA (Minima Controlled Recursive Averaging) d'estimation continue de la Densité Spectrale de Puissance (DSP) du bruit. Cette investigation a concerné également la proposition d'incorporation d'un modèle perceptif, initialement proposé par J.D. Johnston [41], dans le modèle en sous-espaces en association avec le critère VRE et avec proposition d'une nouvelle fonction de suppression optimisée.

Le choix des méthodes de décomposition en sous-espace est justifié par leur immunité contre le phénomène du bruit musical.

De plus, l'utilisation d'un modèle de perception pour la deuxième méthode s'avère très opportune pour améliorer également l'intelligibilité du signal, souvent affecté par la distorsion qu'engendrent ces méthodes sur le signal rehaussé.

I.3. Objectifs de la thèse

L'objectif de cette thèse peut être axé principalement sur la proposition de deux nouvelles méthodes de rehaussement en sous-espace, en l'occurrence KLT-VRE-MCRA [50] et PKLT-VRE [51].

La première méthode KLT-VRE-MCRA proposée est basée sur une Analyse en Composantes Principales (PCA) ou encore désignée par Transformée de Karhunen-Loève (KLT) [52-53], qui consiste à combiner avec la méthode KLT-VRE [54] trois méthodes d'estimation continue de la DSP du bruit MCRA [13], MCRA2 [14] et IMCRA [15] destinées à traquer les variations brusques du spectre du bruit et réduire, à partir d'un lissage temporel, les fluctuations entre segments de présence et d'absence de la parole. La valeur de la DSP du bruit sera ensuite utilisée pour le calcul de la nouvelle matrice du gain de suppression.

La deuxième méthode proposée PKLT-VRE consiste, dans une première étape, à améliorer la méthode de décomposition en sous-espace PKLT [42] qui a été dérivée à partir d'un modèle

perceptif [41], avant de procéder, dans une seconde étape, à sa combinaison avec la méthode KLT-VRE [54]. Une nouvelle fonction de suppression, développée et optimisée par une fonction de régression linéaire, a été proposée.

Dans cette thèse nous nous sommes intéressés au contexte de rehaussement de la parole suivant : (i) Cas mono-capteur ; (ii) Bruit additif non corrélé ; (iii) utilisation, pour des fins de validation des résultats obtenus, de bases de données disjointes.

I.4. Organisation de la thèse

La présente thèse est organisée selon les chapitres suivants :

Dans le chapitre I, nous introduisons en premier lieu le domaine du rehaussement de la parole avons de poursuivre avec les motivations, les objectifs et les contributions se rapportant à ce domaine de recherche.

Dans le chapitre II, nous présentons quelques méthodes standards de rehaussement de la parole, en particulier les méthodes basées sur la soustraction spectrale et sur des modèles statistiques, le filtrage de Wiener, le filtrage de Kalman ainsi que celles fondées sur la décomposition en sous-espaces.

Dans le chapitre III, nous nous intéressons particulièrement aux méthodes de rehaussement en sous-espace ainsi qu'à la présentation du critère de la Variance de l'Erreur de Reconstruction (VRE : Variance Reconstruction Error).

Le chapitre IV, sera consacré à la présentation des deux méthodes proposées dans cette thèse "KLT-VRE-MCRA" et "PKLT-VRE". La robustesse de ces deux méthodes a été validée par le recours à l'évaluation objective de leurs performances et dont les résultats ont été concluants.

Enfin, nous clôturons cette thèse par une conclusion générale en faisant ressortir les éventuelles perspectives devant assurer la continuité de nos travaux de recherche.

Chapitre II

Méthodes de rehaussement de la parole

II.1. Introduction

Le rehaussement de la parole peut être abordé selon deux méthodes différentes dites paramétriques et non paramétriques, basées respectivement sur un modèle mathématique de production de la parole et sur des estimations du bruit. Nous présentons dans ce chapitre quelques méthodes de l'état de l'art les plus utilisées, notamment celles relatives à la Soustraction Spectrale, aux estimateurs ML et MMSE ainsi qu'aux filtrages de Wiener paramétrique et de Kalman. Les différentes méthodes dérivées seront également présentées dans ce chapitre.

II.2. Méthode de Soustraction Spectrale (SS)

La soustraction spectrale consiste à atténuer plus ou moins les composantes spectrales du signal bruité en fonction de l'estimation du niveau de bruit. Nous allons introduire, dans ce chapitre, l'algorithme de Berouti [18], considéré comme une méthode de Soustraction Spectrale de référence.

Soit un signal bruité x composé d'un signal propre s corrompu par un bruit additif n non corrélé. Le signal bruité peut être exprimé par :

$$x(t) = s(t) + n(t) \quad (\text{II.1})$$

Nous désignons par $\hat{s}(t)$ le signal rehaussé qui est une estimation de $s(t)$.

Nous supposons que le signal de parole est quasi-stationnaire sur des fenêtres d'analyse de 20 à 30ms. La Transformée de Fourier à Court Terme (TFCT) de $x(t)$ (II.1) est donnée par :

$$X(\omega) = S(\omega) + N(\omega) \quad (\text{II.2})$$

où $X(\omega)$, $S(\omega)$ et $N(\omega)$ désignent respectivement le spectre de $x(t)$, $s(t)$ et $n(t)$.

$X(\omega)$ peut être exprimée en forme polaire comme suit :

$$X(\omega) = |X(\omega)|e^{j\Phi_x(\omega)} \quad (\text{II.3})$$

$|X(\omega)|$ et $\Phi_x(\omega)$ représentent respectivement l'amplitude et la phase de $X(\omega)$.

On peut également exprimer le bruit par $N(\omega) = |N(\omega)|e^{j\Phi_n(\omega)}$. L'amplitude du bruit $|N(\omega)|$ est inconnue, ce qui nous conduit donc à la remplacer par sa valeur moyenne, estimée durant la période d'absence d'activité vocale.

Dans les méthodes de soustraction spectrale, il a été supposé que le bruit n'affecte pas la phase du signal. Pour cela, on considère uniquement l'amplitude spectrale à court terme du bruit. La valeur estimée de $\Phi_n(\omega)$ sera donc remplacée par $\Phi_x(\omega)$.

En remplaçant $X(\omega)$ et $N(\omega)$ dans (II.2), la valeur estimée du spectre du signal propre est donnée par :

$$\hat{S}(\omega) = [|X(\omega)| - |\hat{N}(\omega)|]e^{j\Phi_x(\omega)} \quad (\text{II.4})$$

Où $|\hat{N}(\omega)|$ est la valeur estimée du spectre d'amplitude du bruit calculée durant la période d'absence d'activité vocale. Le signal rehaussé peut être obtenu en calculant la transformée de Fourier inverse de $\hat{S}(\omega)$. On peut remarquer à partir de (II.4) que $|\hat{S}(\omega)| = |X(\omega)| - |\hat{N}(\omega)|$ peut prendre des valeurs négatives en raison des erreurs d'estimation du bruit. Plusieurs solutions sont envisageables pour remédier à ce problème [18-19]. Parmi ces techniques on peut citer la rectification demi-onde qui permet de mettre à zéro ces valeurs négatives, comme suit :

$$\begin{cases} |\hat{S}(\omega)| = |X(\omega)| - |\hat{N}(\omega)| & \text{si } |X(\omega)| > |\hat{N}(\omega)| \\ 0 & \text{ailleurs} \end{cases} \quad (\text{II.5})$$

L'algorithme de soustraction spectrale en amplitude (II.4) peut être étendu à la soustraction spectrale de puissance en multipliant $X(\omega)$ (II.2) par sa conjuguée $X^*(\omega)$ comme suit :

$$\begin{aligned} |X(\omega)|^2 &= |S(\omega)|^2 + |N(\omega)|^2 + S(\omega).N^*(\omega) + S^*(\omega).N(\omega) \\ &= |S(\omega)|^2 + |N(\omega)|^2 + 2.\text{Re}\{S(\omega).N^*(\omega)\} \end{aligned} \quad (\text{II.6})$$

Les termes $|N(\omega)|^2$, $S(\omega).N^*(\omega)$ et $S^*(\omega).N(\omega)$ ne peuvent pas être obtenus directement, mais approximatés par $E\{|N(\omega)|^2\}$, $E\{S(\omega).N^*(\omega)\}$ et $E\{S^*(\omega).N(\omega)\}$, où $E\{\cdot\}$ désigne l'opérateur

espérance mathématique. Généralement $E\{|N(\omega)|^2\}$ est estimée durant la période d'absence d'activité vocale est désignée par $|\hat{N}(\omega)|^2$.

Si nous supposons que $n(t)$ et $s(t)$ est à moyenne nulle et non corrélé avec $s(t)$, les termes $E\{S(\omega).N^*(\omega)\}$ et $E\{S^*(\omega).N(\omega)\}$ deviennent nuls. La valeur estimée du spectre de puissance du signal propre devient :

$$|\hat{S}(\omega)|^2 = |X(\omega)|^2 - |\hat{N}(\omega)|^2 \quad (\text{II.7})$$

L'équation (II.7) décrit l'algorithme de soustraction spectrale en puissance et peut être exprimée par :

$$|\hat{S}(\omega)|^2 = H^2(\omega)|X(\omega)|^2 \quad (\text{II.8})$$

où

$$H(\omega) = \sqrt{1 - \frac{|\hat{N}(\omega)|^2}{|X(\omega)|^2}} \quad (\text{II.9})$$

$H(\omega)$ désigne le gain de suppression ou la fonction de suppression.

II.3. Filtrage de Wiener Paramétrique

Les techniques de rehaussement de la parole les plus couramment utilisées sont des généralisations de techniques comme le filtrage de Wiener à court-terme [20] et les méthodes de soustraction spectrale. Ces méthodes ont été largement étudiées et employées dans plusieurs applications. Cette famille d'algorithmes peut être rassemblée sous une même catégorie appelée filtrage de Wiener paramétrique ou soustraction spectrale paramétrique pouvant être exprimés par une formulation paramétrique générale. Le gain spectral correspondant à cette paramétrisation est alors défini par :

$$H(\omega) = \left[1 - \left(\frac{|\hat{N}(\omega)|^2}{|X(\omega)|^2} \right)^{\frac{\delta-1}{2}} \right]^{\rho} \quad (\text{II.10})$$

Où δ et ρ sont deux paramètres fixes. A partir de cette généralisation et selon les différentes valeurs prises par le couple (δ, ρ) , indiquées dans le tableau II.1, on retrouve les différentes

méthodes de réduction du bruit qui sont la soustraction spectrale en puissance, en amplitude et le filtrage de Wiener à court terme.

Méthode	δ	ρ
Soustraction spectrale en puissance	2	1/2
Soustraction spectrale en amplitude	1	1
Filtrage de Wiener à court-terme	2	1

Tableau II. 1 : filtrage de Wiener paramétrique et méthodes correspondantes pour des valeurs de (δ, ρ) .

Considérons la variable $R_{post}(\omega)$ définie par :

$$R_{post}(\omega) = \frac{|X(\omega)|^2}{|\hat{N}(\omega)|^2} \quad (\text{II.11})$$

Sous les hypothèses de décorrélation du signal source et du bruit, $R_{post}(\omega)$ peut être considérée comme une estimation du Rapport Signal sur Bruit (RSB) instantané. La règle de suppression de l'équation (II.10) devient :

$$H(\omega) = \left[1 - (R_{post}(\omega))^{-\frac{\delta}{2}} \right]^\rho \quad (\text{II.12})$$

Cette reformulation de la règle de suppression pour les méthodes de filtrage de Wiener paramétrique montre que le gain de suppression ne dépend que du RSB instantané, exprimé par la variable $R_{post}(\omega)$. La règle de suppression de l'équation (II.12) est donc une règle qui ne se définit qu'à partir d'une information instantanée.

II.4. Phénomène du bruit musical

En utilisant des valeurs de (δ, ρ) issues du tableau II.1, les algorithmes de filtrage de Wiener paramétrique parviennent à réduire de manière très efficace le niveau du bruit de fond. Néanmoins, le bruit résiduel qui subsiste dans le signal rehaussé après modification spectrale est très gênant du point de vue perceptif. Ce phénomène est une caractéristique des méthodes de soustraction spectrale à court terme. Du fait du caractère tonal particulier de cet artéfact, il est désigné par le terme de bruit musical. Son spectre à court terme correspond approximativement à une distribution aléatoire de pics spectraux. L'origine du bruit musical

est la variance des estimateurs locaux de la densité spectrale des signaux [1], [18-19]. En effet, comme le spectre à court terme du bruit fluctue autour de sa valeur moyenne, son amplitude atteint à certains instants et pour certains indices fréquentiels des valeurs largement supérieures à la moyenne.

Considérons le cas où le signal est fortement bruité dans une certaine zone du spectre où le RSB est relativement bas. Sporadiquement, du fait de la variance du bruit, l'amplitude spectrale du signal atteint des valeurs largement supérieures au niveau moyen estimé du bruit conduisant donc à une surestimation locale et instantanée du RSB. Selon la règle de suppression définie dans l'équation (II.12), la bande de fréquences correspondante est alors traitée comme du signal utile et elle est relativement moins atténuée que les composantes fréquentielles voisines. De manière sporadique, des pics fréquentiels isolés se dégagent donc du spectre atténué, engendrant ainsi le phénomène de bruit musical.

Plusieurs méthodes ont été proposées pour supprimer le bruit musical [55-59]. Cependant, il est difficile de réduire ce dernier sans pour autant affecter le signal de parole. En général, les différentes approches proposées considèrent des compromis entre la réduction du bruit et le niveau des distorsions pouvant affecter le signal rehaussé.

A titre d'exemple, la figure II.1 donne une représentation temporelle la phrase "*Hedge apples may stain your hands green*", prélevée de la base de données NOIZEUS [60]. Cette phrase est corrompue par un bruit de chahut (babble noise) à 0 dB de RSB. A partir du signal rehaussé, on peut tirer les remarques suivantes : (i) représente une atténuation du signal rehaussé, due essentiellement à une surestimation entraînant sa distorsion (ii) représente une sous-estimation du bruit qui a engendré la présence de ce dernier dans une partie du signal rehaussé.

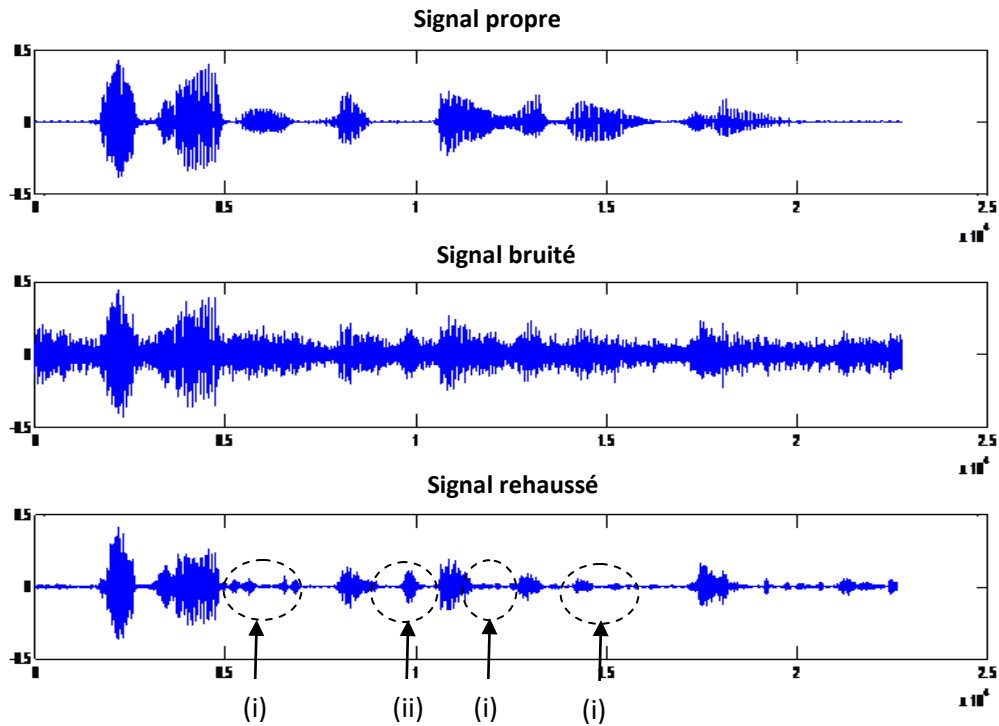


Figure II. 1 : Représentation temporelle d'un signal propre, corrompu par un bruit de chahut à 0B de RSB et rehaussé par la méthode de Soustraction Spectrale de Berouti [18].

Berouti et al. ont proposé une méthode de minimisation du bruit musical qui consiste à réaliser un compromis entre une sous-estimation et surestimation du niveau de bruit. Cette méthode consiste à soustraire une surestimation du spectre de puissance du bruit, tout en empêchant les composantes spectrales résultantes de descendre en dessous d'une valeur minimale prééglée. Cette méthode est donnée comme suit :

$$|\hat{S}(\omega)|^2 = \begin{cases} |X(\omega)|^2 - \alpha |\hat{N}(\omega)|^2 & \text{si } |X(\omega)|^2 > (\alpha + \beta) |\hat{N}(\omega)|^2 \\ \beta |\hat{N}(\omega)|^2 & \text{autrement} \end{cases} \quad (\text{II.13})$$

où α ($\alpha \geq 1$) est le facteur de sur-soustraction et β ($0 < \beta \ll 1$) le spectral floor (paramètre d'aplatissement spectral). En effet, l'opération de soustraction de la valeur estimée du spectre du bruit à partir du spectre de la parole bruité engendre l'apparition de pics dans le spectre. Certains de ces pics sont à large bande, tandis que d'autres sont à bande étroite qui se présentent comme des pointes dans le spectre. La sur-soustraction du bruit, avec $\alpha > 1$,

réduira considérablement l'amplitude des pics large bande. Cette opération est suivie d'un aplatissage spectral pour éliminer davantage les résidus restants.

La fonction de suppression est donnée par :

$$H(\omega) = \left(\frac{|\gamma(\omega)| - \alpha}{|\gamma(\omega)|} \right)^{\frac{1}{2}} \quad (\text{II.14})$$

avec $\gamma(\omega) = R_{post}(\omega)$ est le $R_{post}(\omega)$

La figure II. 2, montre le tracé de $H(\omega)$ en fonction de $\gamma(\omega)$ pour différentes valeurs de α . L'atténuation est beaucoup plus importante que lorsque α est plus grand. Par exemple pour $\gamma(\omega) = 8\text{dB}$, l'atténuation obtenue pour $\alpha = 1$ et -0.74dB , avec $\alpha = 3$, elle est de -2.8dB et avec $\alpha = 5$, elle de -6.8dB . On peut aussi remarquer quand l'aplanissement spectral apparait. Par exemple si $\alpha = 5$, l'aplanissement spectral apparait lorsque $R_{post}(\omega)$ est inférieur à 5dB , et lorsque $\alpha = 1$, il apparait lorsque $R_{post}(\omega)$ est inférieur à 0dB .

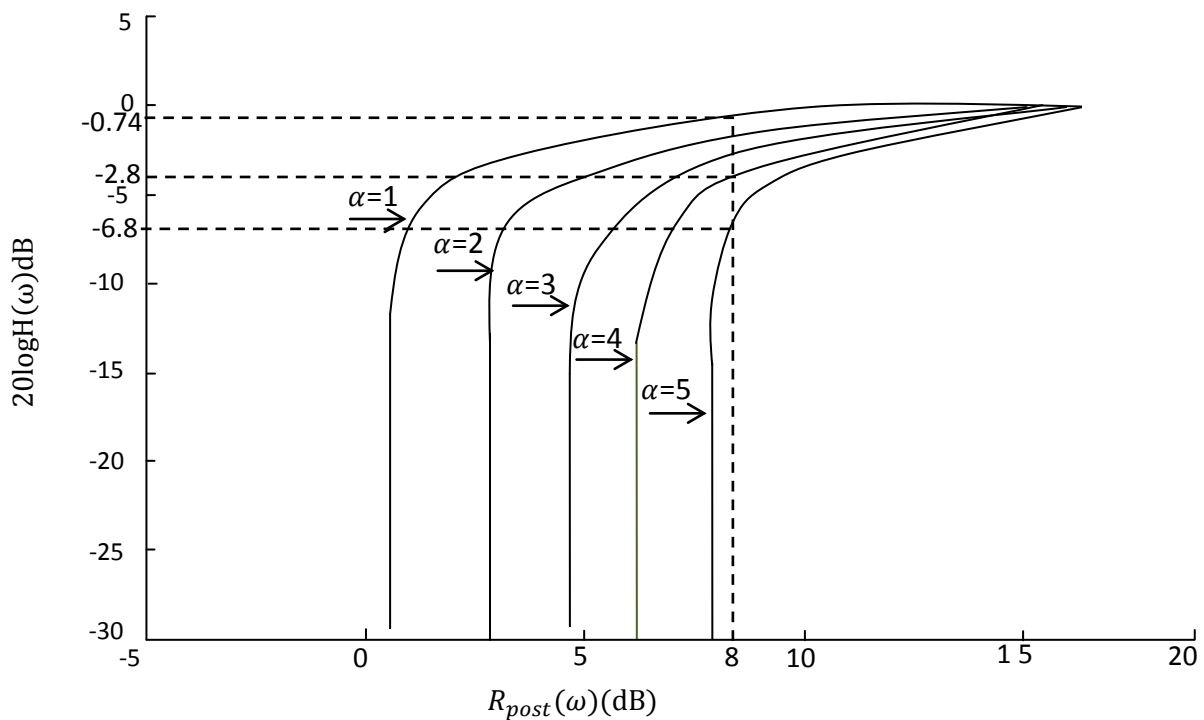


Figure II. 2 : Courbes d'atténuations de l'algorithme de Soustraction Spectrale pour différentes valeurs de α [1]

II.5. Techniques de Soustraction Spectrale

Nous avons vu précédemment que le choix des paramètres α et β influence plus ou moins les performances des algorithmes de Soustraction Spectrale. En revanche, cette méthode SS de

référence a été améliorée par des méthodes qui proposent de combiner différemment ces paramètres, comme suit :

II.5.1. Technique de Soustraction Spectrale Non Linéaire (NSS-SB)

Dans [18], il a été supposé que le bruit affecte de la même façon toutes les composantes spectrales. Pour cela, un seul facteur de soustraction α est utilisé pour soustraire une surestimation du bruit sur tout le spectre. Par ailleurs, cette supposition n'est plus opportune dans le cas des bruits environnementaux tel que les bruits de car, de restaurant, de rue, etc. Certains bruits affectent plus la région des basses fréquences que la région des hautes fréquences, d'où la nécessité de favoriser l'utilisation d'un facteur de soustraction dépendant la fréquence.

La méthode NSS-SB (Non linear Spectral Subtraction) [61] peut être considérée comme une modification de la méthode de Berouti, été ce, par l'utilisation d'un facteur de soustraction variable associé à un processus de soustraction non linéaire.

Le filtre de suppression est donné par :

$$H_{NSS}(\omega) = \frac{|\bar{X}(\omega)| - \alpha(\omega)\tilde{N}(\omega)}{|\bar{X}(\omega)|} \quad (\text{II.15})$$

- $\alpha(\omega)$ est le facteur de soustraction dépendant de la fréquence et du RSB a posteriori :

$$\alpha(\omega) = \frac{1}{1+Y\rho(\omega)} \quad (\text{II.16})$$

Y est un scalaire et $\rho(\omega) = \frac{|\bar{X}(\omega)|}{|\bar{N}(\omega)|}$ la racine carrée du RSB a posteriori

- $\tilde{N}(\omega)$ une fonction non linéaire du spectre du bruit, obtenue à partir de la valeur maximale du spectre d'amplitude du bruit $|\hat{N}_i(\omega)|$ sur les 40 trames précédentes :

$$\tilde{N}(\omega) = \max_{i-40 \leq j \leq i} (|\hat{N}_j(\omega)|) \quad (\text{II.17})$$

II.5.2. Technique de Soustraction Spectrale Multi-Bandes (MB-SB)

L'idée de base est similaire à celle donnée par la méthode NSS. En effet, le bruit n'affecte pas uniformément le signal de parole. La principale différence entre l'approche MB-SB (MultiBand spectral subtraction) [62] et l'approche NSS réside dans l'estimation du facteur de sur-soustraction. Dans MB, le facteur de soustraction est estimé pour chaque bande de fréquence, tandis que dans NSS, il est estimé pour chaque fréquence. L'inconvénient de

l'estimation d'un seul facteur de soustraction pour chaque fréquence et que le RSB de chaque fréquence peut radicalement changer d'une trame à une autre. Ces changements radicaux sont en partie responsables de la distorsion de la parole (bruit musical) rencontrée dans la soustraction spectrale. Le spectre du signal bruité est divisé en L bandes non recouvrantes et la soustraction du bruit est réalisée indépendamment dans chaque bande. Ce processus peut être réalisé soit dans le domaine temporel par le biais de filtres passe-bande ou dans le domaine fréquentiel à l'aide des fenêtres appropriées. Le filtre de suppression est donné par :

$$H_{MB}(\omega_k) = \frac{|\bar{X}_i(\omega_k)|^2 - \alpha_i \delta_i |\hat{N}_i(\omega)|^2}{|\bar{X}_i(\omega_k)|^2} \quad (\text{II.18})$$

avec $b_i < \omega_k = 2\pi k/L < e_i$ ($k = 0, 1, \dots, L - 1$) a fréquence discrète, b_i et e_i sont les fréquences de début et de fin, $|\hat{N}_i(\omega)|^2$ est la valeur estimée du spectre de puissance du bruit (obtenue et mise à jour durant les segments d'absence d'activité vocale), α_i est le facteur de soustraction et δ_i est un facteur de soustraction supplémentaire utilisé pour ajuster individuellement chaque bande afin d'adapter le processus de soustraction.

Le schéma bloc de l'algorithme de soustraction spectrale MB est donné sur la figure II. 3.

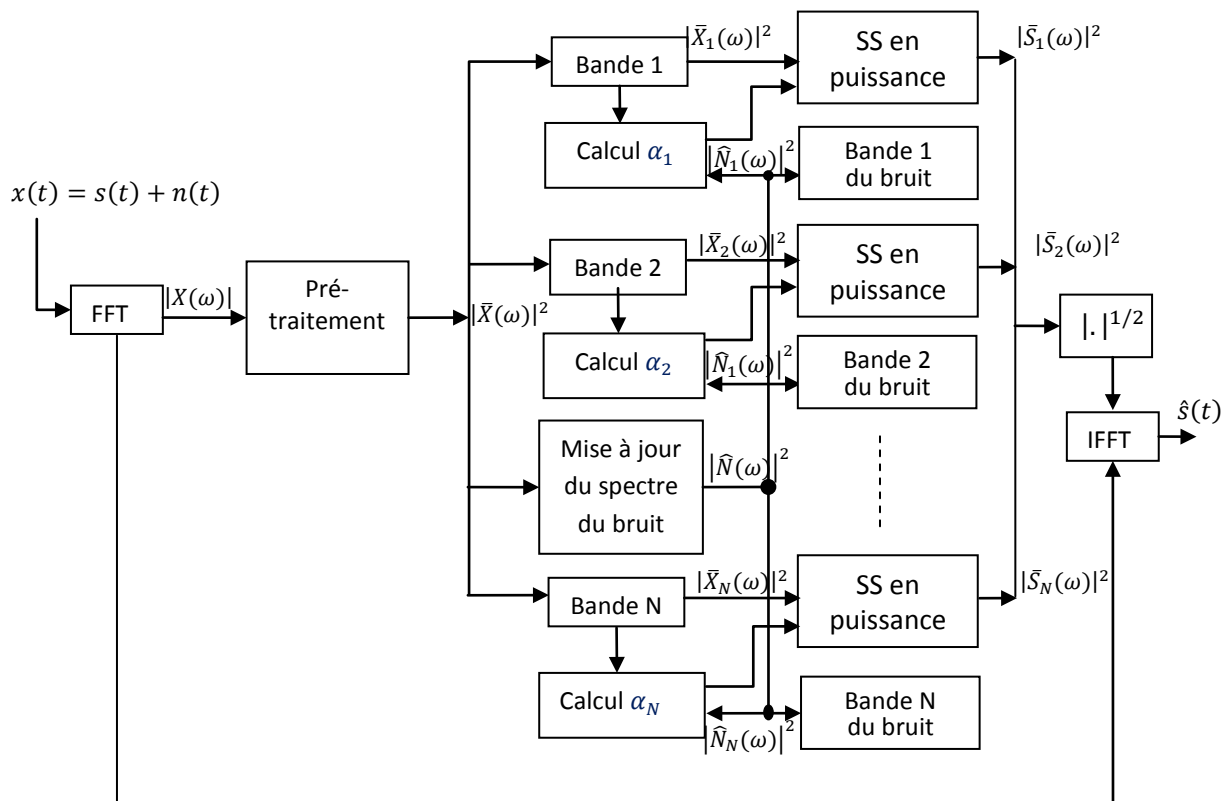


Figure II. 3 : Schéma bloc de l'algorithme de soustraction spectrale MB [62]

II.5.3. Technique de Soustraction Spectrale basée sur le critère MMSE

Dans les méthodes NSS et MB, les paramètres α et β ont été déterminés expérimentalement. Dans la méthode de Soustraction Spectrale basée sur MMSE (MMSE Spectral Subtraction Algorithm) [63] la sélection desdits paramètres s'effectue de manière optimale, au sens de l'erreur quadratique moyenne. Le filtre de suppression est donné par :

$$H_{MMSE}(\omega) = \left\{ \frac{\xi^p(\omega)}{\delta_p + \xi^p(\omega)} \left[\frac{|X(\omega)|^p - |\hat{N}(\omega)|^p}{|X(\omega)|^p} \right] \right\}^{1/p} \quad (\text{II.19})$$

où p est un exposant de puissance désignant la forme générale de l'algorithme de soustraction spectrale. δ_p est une constante ($\delta_p = 0.2146, 0.5$, et 0.7055 respectivement pour $p = 1, 2$ et 3).

$\xi(\omega) = \frac{E\{|S(\omega)|^2\}}{E\{|N(\omega)|^2\}}$, le RSB a priori ne peut pas être effectué directement puisque $S(\omega)$ est inconnu. Cependant, $\xi(\omega)$ a été approximé de la même manière que dans la méthode statistique basée sur le critère MMSE (II.6.3).

II.5.4. Technique de Soustraction Spectrale basée sur le Moyennage Adaptative du Gain (AGA)

Nous savons que les deux facteurs responsables de la présence du bruit musical après soustraction spectrale sont la grande variance de l'estimation spectrale et la variabilité. Pour cela, Gustafsson et al [64] ont proposé une méthode de Soustraction Spectrale dite AGA (Adaptative Gain Averaging) qui consiste à diviser la trame d'analyse courante en sous trames en vue d'obtenir un spectre à faible résolution. Le spectre d'amplitude de chaque sous trame est moyenné pour obtenir un spectre à faible variance suivi d'un lissage temporel de la fonction de suppression par le biais d'un moyennage exponentiel adaptatif.

Le filtre de suppression est donné par :

$$H_i^{(M)}(\omega) = 1 - \alpha \left| \frac{\hat{N}_i^{(M)}(\omega)}{\bar{X}_i^{(M)}(\omega)} \right| \quad (\text{II.20})$$

Où α est le facteur de soustraction et $\hat{N}_i^{(M)}(\omega)$ est la valeur estimée du spectre d'amplitude du bruit, mis à jour durant les segments d'absence d'activité vocale. Pour réduire la variabilité de la fonction de suppression, $H_i^{(M)}(\omega)$ est moyennée selon la relation suivante :

$$\bar{H}_i^{(M)}(\omega) = a_i \bar{H}_{i-1}^{(M)}(\omega) + (1 - a_i) H_i^{(M)}(\omega) \quad (\text{II.21})$$

$\bar{H}_i^{(M)}(\omega)$ désigne la fonction de suppression lissée et a_i le paramètre de lissage adaptatif.

La figure II.4, montre les spectrogrammes du signal propre de la phrase "A good book informs of what we ought to know" des signaux propres, bruités ainsi que ceux rehaussés par les méthodes de SS_Berouti, SS_MB et SS-AGA. Le type de bruit utilisé est le chahut (babble) à 5dB de RSB d'entrée que nous avons prélevé à partir de la base de données NOIZEUS.

A partir de ces spectrogrammes, nous pouvons remarquer que la qualité et l'intelligibilité du signal est conditionnée par le meilleur compromis à réaliser entre la réduction du bruit et la distorsion du signal.

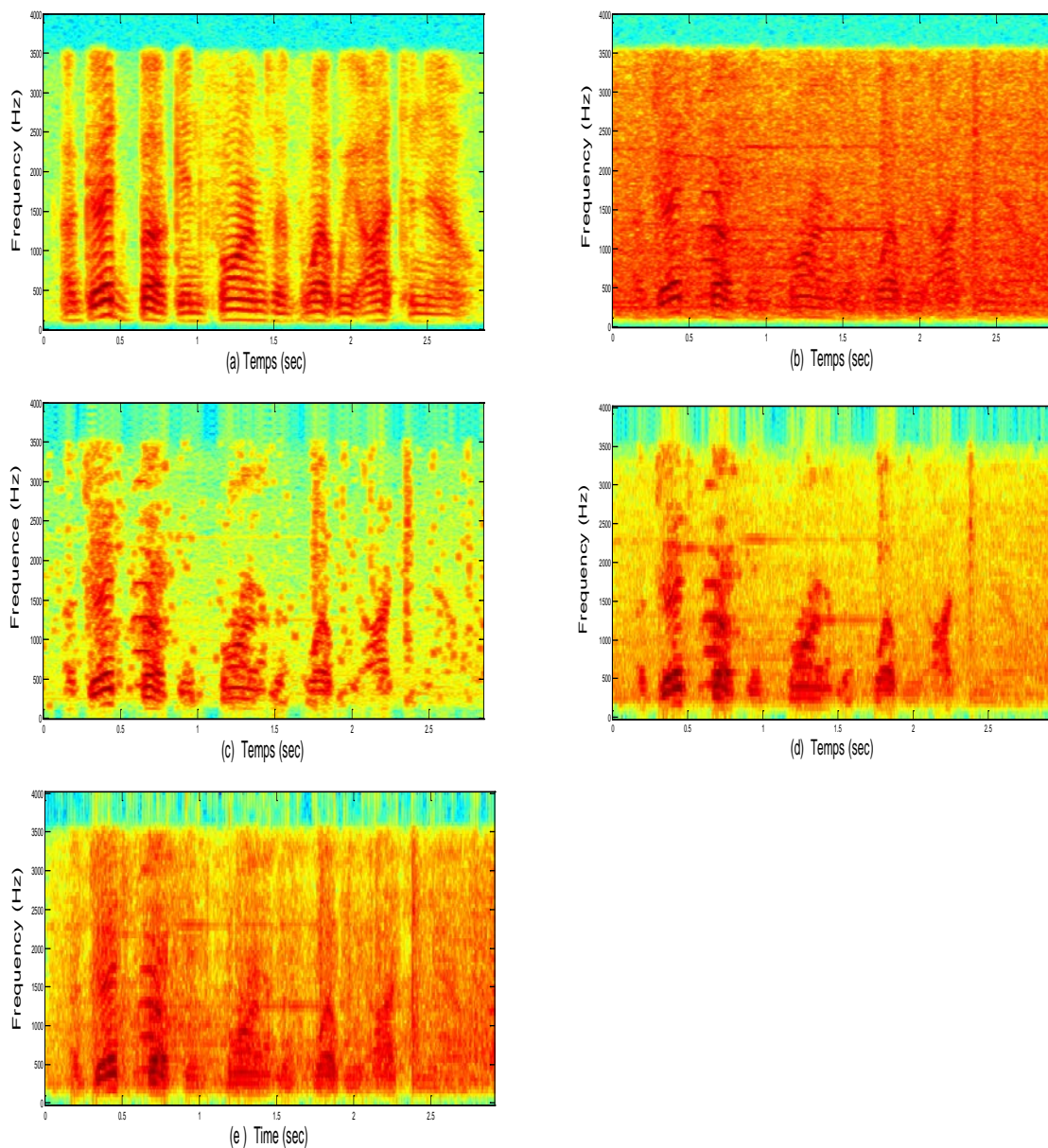


Figure II. 4 : Spectrogrammes d'une phrase issue de la base de données NOIZEUS avec (a) signal propre (b) signal bruité avec du chahut à 5dB, (c) signal rehaussé par SS_Berouti (d) signal rehaussé par SS_MB et (e) signal rehaussé par SS_AGA.

II.6. Méthodes de rehaussement de la parole basées sur des modèles statistiques

Le rehaussement de la parole peut être considéré comme un problème d'estimation statistique. Étant donné un ensemble de mesures dépendant d'un paramètre inconnu, on cherche donc à trouver un estimateur linéaire de ce paramètre. Dans notre cas, les mesures correspondent aux coefficients de la Transformée de Fourier Discrète (TFD) du signal bruité et les paramètres recherchés sont l'ensemble des coefficients TFD du signal propre. Il existe plusieurs techniques issues de la théorie de l'estimation [65-66] qui sont utilisées pour calculer ces estimateurs non linéaires. Parmi ces techniques, on peut citer les estimateurs Bayésiens et les estimateurs du Maximum de Vraisemblance (ML). Les différences entre ces estimateurs résident principalement dans les hypothèses émises sur les paramètres inconnus (déterministe, mais inconnu, aléatoire, etc.) et sur la forme du critère d'optimisation utilisé.

II.6.1. Estimateurs du Maximum de Vraisemblance (ML)

Dans le domaine du rehaussement de la parole, cette approche a été initialement proposée par Mc Aulay et al. [21].

Nous supposons que nous disposons d'un ensemble de données $\mathbf{x} = [x(0), x(1), \dots, x(L-1)]$ dépendant d'un paramètre inconnu θ . Dans le domaine du rehaussement de la parole \mathbf{x} (ensemble des données observées) représente le spectre d'amplitude du signal bruité et le paramètre θ représente le spectre d'amplitude du signal propre (inconnu). En outre, nous supposons que la Fonction Densité de Probabilité (FDP) de \mathbf{x} soit connue et donnée par $p(\mathbf{x}; \theta)$. Tant que le paramètre θ influe sur la probabilité de \mathbf{x} , nous devrions être en mesure de déduire les valeurs de θ à partir des valeurs observées de \mathbf{x} . La valeur de θ la plus probable qui permet de produire \mathbf{x} et celle qui maximise $p(\mathbf{x}, \theta)$, comme suit :

$$\hat{\theta}_{ML} = \arg \max_{\theta} p(\mathbf{x}; \theta) \quad (\text{II.22})$$

$\hat{\theta}_{ML}$ est désignée par la valeur estimée de θ , au sens du maximum de vraisemblance, $p(\mathbf{x}, \theta)$ est la fonction de vraisemblance. $\hat{\theta}_{ML}$ peut être calculée en dérivant, par rapport à θ , la fonction de vraisemblance $p(\mathbf{x}, \theta)$ et en déduisant la valeur de θ qui permet d'annuler $p(\mathbf{x}, \theta)$.

Le spectre de la version échantillonnée du signal bruité de l'équation (II.1) peut être donné comme suit :

$$X(\omega_k) = S(\omega_k) + N(\omega_k) \quad (\text{II.23})$$

avec $\omega_k = 2\pi k/L$ et $k = 0,1,2,\dots,L-1$, ou L représente la longueur de la trame en échantillons. L'équation (II.23) peut être exprimée en forme polaire comme suit :

$$X_k e^{j\theta_x(k)} = S_k e^{j\theta_s(k)} + N_k e^{j\theta_b(k)} \quad (\text{II.24})$$

où $[X_k, S_k, N_k]$ et $[\theta_x(k), \theta_s(k), \theta_n(k)]$ indiquent respectivement les amplitudes et les phases du signal bruité, du signal propre et du bruit.

McAulay et al. ont considéré que X_k et $\theta_x(k)$ sont inconnus et déterministes. La FDP du bruit $N(\omega_k)$ est considérée comme gaussienne à moyenne nulle. Les parties réelles et imaginaires de $N(\omega_k)$ sont à variance $\sigma_b(k)/2$. La FDP de $X(\omega_k)$ est aussi une gaussienne avec variance $\sigma_n(k)$ et moyenne $S_k e^{j\theta_s(k)}$:

$$\begin{aligned} p(X(\omega_k); S_k, \theta_s(k)) &= \frac{1}{\pi\sigma_n(k)} \exp \left[-\frac{|X(\omega_k) - S_k e^{j\theta_s(k)}|^2}{\sigma_n(k)} \right] \\ &= \frac{1}{\pi\sigma_n(k)} \exp \left[-\frac{X_k^2 - 2S_k \operatorname{Re}\{e^{j\theta_s(k)} X(\omega_k)\} + S_k^2}{\sigma_n(k)} \right] \end{aligned} \quad (\text{II.25})$$

Pour obtenir l'estimateur du maximum de vraisemblance de S_k , nous devons calculer le maximum de $p(X(\omega_k); S_k, \theta_s(k))$, qui est une tâche difficile du fait que $p(X(\omega_k); S_k, \theta_s(k))$ est une fonction de deux paramètres inconnus : l'amplitude et la phase. La phase peut être éliminée en maximisant la fonction de vraisemblance moyenne suivante :

$$p_L(X(\omega_k); S_k) = \int_0^{2\pi} p(X(\omega_k); S_k, \theta_s(k)) p(\theta_s) d\theta_s \quad (\text{II.26})$$

Supposons que θ_s est uniformément distribuée sur $[0, 2\pi]$, ce qui implique que $p(\theta_s) = 1/2\pi$ pour $\theta_s \in [0, 2\pi]$. La fonction de vraisemblance devient :

$$p_L(X(\omega_k); S_k) = \frac{1}{\pi\sigma_n(k)} \exp \left[-\frac{X_k^2 + S_k^2}{\sigma_n(k)} \right] \frac{1}{2\pi} \int_0^{2\pi} \exp \left[\frac{2S_k \operatorname{Re}\{e^{-j\theta_s} X(\omega_k)\}}{\sigma_n(k)} \right] d\theta_s \quad (\text{II.27})$$

L'intégral de l'équation (II.27) représente la fonction de Bessel modifiée du premier ordre, qui peut être approximée par :

$$p_L(X(\omega_k); S_k) = \frac{1}{\pi\sigma_b(k)} \left(2\pi \frac{2S_k X_k}{\sigma_n(k)} \right)^{-1/2} \exp \left[-\frac{X_k^2 + S_k^2 - 2X_k S_k}{\sigma_n(k)} \right] \quad (\text{II.28})$$

Après dérivation par rapport à S_k et annulation du logarithme de la fonction de vraisemblance $\log p_L(X(\omega_k); S_k)$, nous obtenons la valeur estimée du maximum de vraisemblance du spectre d'amplitude comme suit :

$$\hat{S}_k = \frac{1}{2} \left[X_k + \sqrt{X_k^2 - \sigma_n(k)} \right] \quad (\text{II.29})$$

On peut exprimer la valeur estimée du spectre du signal original, en remplaçant θ_s par θ_x comme suit:

$$\begin{aligned} \hat{S}(\omega_k) &= \hat{S}_k e^{j\theta_x} = \hat{S}_k \frac{X(\omega_k)}{X_k} \\ &= \left[\frac{1}{2} + \frac{1}{2} \sqrt{\frac{X_k^2 - \sigma_n(k)}{X_k^2}} \right] X(\omega_k) \end{aligned} \quad (\text{II.30})$$

On appelle $Y_k = X_k^2 / \sigma_n(k)$ le RSB a posteriori, calculé à partir du signal bruité. L'équation (II.30) peut être exprimée par :

$$\begin{aligned} \hat{S}(\omega_k) &= \left[\frac{1}{2} + \frac{1}{2} \sqrt{\frac{Y_k - 1}{Y_k}} \right] X(\omega_k) \\ &= H_{ML}(Y_k) X(\omega_k) \end{aligned} \quad (\text{II.31})$$

$H_{ML}(Y_k)$ désigne la fonction de suppression de l'estimateur ML.

On peut obtenir à partir de l'estimation du maximum de vraisemblance, une fonction de suppression différente en supposant que le signal bruité et le bruit peuvent être modélisés comme des processus aléatoires gaussiens indépendants avec $\sigma_s(k)$ déterministe et inconnue. Supposons que le signal propre et le bruit sont indépendants, la variance du signal bruité $\sigma_x(k)$ peut être exprimée par : $\sigma_x(k) = \sigma_s(k) + \sigma_n(k)$ et la densité de probabilité de $X(\omega_k)$ est donnée par :

$$p(X(\omega_k); \sigma_s(k)) = \frac{1}{\pi[\sigma_s(k)+\sigma_n(k)]} \exp \left[-\frac{X_k^2}{\sigma_s(k)+\sigma_n(k)} \right] \quad (\text{II.32})$$

En maximisant la fonction de vraisemblance de $p(X(\omega_k); \sigma_s(k))$ par rapport à $\sigma_s(k)$, nous obtenons :

$$\hat{\sigma}_s(k) = X_k^2 - \sigma_n(k) \quad (\text{II.33})$$

En supposant que $S_k^2 \approx \sigma_s(k)$ et $B_k^2 \approx \sigma_n(k)$ (et $X_k^2 - \sigma_n(k) > 0$), nous obtenons une estimation du spectre d'amplitude du signal comme suit:

$$\hat{S}_k = \sqrt{X_k^2 - N_k^2} \quad (\text{II.34})$$

Il est à noter que cet estimateur de S_k n'est autre que l'estimateur SS en puissance [18]. Pour cela, l'estimateur SS en puissance peut être calculé à travers l'estimateur du maximum de vraisemblance, sachant que le signal bruité et le bruit sont modélisés comme des processus aléatoires indépendants et que la variance du signal original $\sigma_s(k)$ est inconnue mais déterministe. On peut calculer l'estimée du spectre du signal propre comme suit :

$$\begin{aligned} \hat{S}(\omega_k) &= \hat{S}_k e^{j\theta_x} = \hat{S}_k \frac{X(\omega_k)}{X_k} \\ &= \sqrt{\frac{X_k^2 - \sigma_n(k)}{X_k^2}} X(\omega_k) \end{aligned} \quad (\text{II.35})$$

En fonction de Y_k (RSB a posteriori), $\hat{S}(\omega_k)$ devient :

$$\begin{aligned} \hat{S}(\omega_k) &= \sqrt{\frac{Y_k - 1}{Y_k}} X(\omega_k) \\ &= H_{PS}(Y_k) X(\omega_k) \end{aligned} \quad (\text{II.36})$$

$H_{PS}(Y_k)$ représente la fonction de suppression de la méthode de soustraction spectrale en puissance. Il est à noter que si on substitue l'estimée du maximum de vraisemblance de $\sigma_n(k)$ (II.33) dans l'équation du filtre de Wiener [1], on aura l'équation suivante :

$$\hat{S}(\omega_k) = \frac{\sigma_s(k)}{\sigma_s(k)+\sigma_n(k)} X(\omega_k) \quad (\text{II.37})$$

$\hat{S}(\omega_k)$ devient :

$$\begin{aligned}\hat{S}(\omega_k) &= \frac{X_k^2 - \sigma_n(k)}{X_k^2} X(\omega_k) \\ &= \frac{Y_k - 1}{Y_k} X(\omega_k) \\ &= H_{PS}^2(Y_k) X(\omega_k)\end{aligned}\tag{II.38}$$

A partir des équations (II.36) et (II.38), nous remarquons que l'estimateur de Wiener est égal au carré de l'estimateur SS en puissance.

Dans [21], il a été mis en association avec la méthode ML un modèle à deux états incorporant l'incertitude de présence et d'absence du signal de la parole en vue d'améliorer cette méthode. La figure II. 5, montre que la règle de suppression ML assure une atténuation beaucoup plus faible comparée avec les méthodes de Wiener ou SS en puissance.

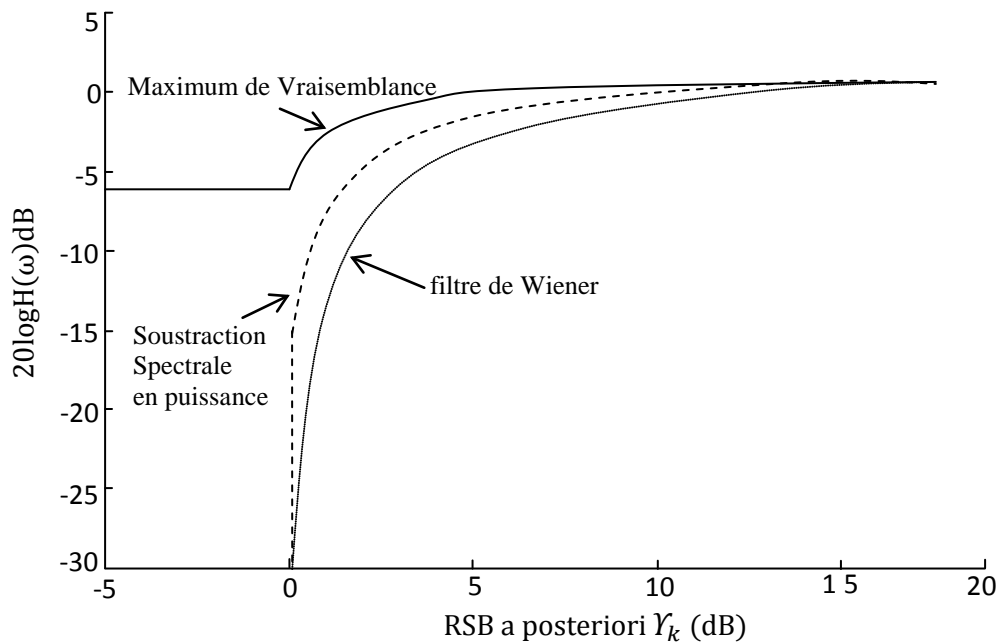


Figure II. 5 : Courbes d'atténuations de la méthode du maximum de vraisemblance [1]

II.6.2. Estimateurs Bayésien

Nous avons supposé dans la section précédente que le paramètre θ est inconnu mais déterministe. Dans cette partie, nous supposons que θ est une variable aléatoire, dont nous devons estimer sa réalisation. Cette approche est appelée approche Bayésienne. Ces estimateurs sont plus performants que les estimateurs ML du fait qu'ils utilisent des connaissances a priori sur θ (càd $p(\theta)$ connue).

II.6.3. Estimateur de l'Erreur Quadratique Moyenne (MSE)

Compte tenu de l'importance que revêt l'utilisation du Spectre d'Amplitude à Court Terme (STSA : Short Term Spectral Amplitude) dans le cadre de l'amélioration de l'intelligibilité et de la qualité de la parole, plusieurs chercheurs ont proposé des méthodes optimales pour obtenir le spectre d'amplitude à partir du signal bruité, notamment celles permettant de minimiser l'erreur quadratique moyenne entre l'amplitude du signal estimée et propre comme suit :

$$e = E \{ (\hat{S}_k - S_k)^2 \} \quad (\text{II.39})$$

Où \hat{S}_k et S_k représentent respectivement la valeur estimée du spectre d'amplitude et le spectre d'amplitude du signal propre à la fréquence ω_k .

Dans une approche MSE, l'espérance mathématique de l'équation (II.39) est calculée par rapport à la FDP conjointe $p(X, S_k)$ comme suit :

$$B_{mse}(\hat{S}_k) = \iint (S_k - \hat{S}_k)^2 p(X, S_k) dX dS_k \quad (\text{II.40})$$

La minimisation de l'équation (II.40) par rapport à \hat{S}_k conduit à l'estimateur MMSE [22], donné par:

$$\begin{aligned} \hat{S}_k &= \int S_k p(S_k|X) dS_k \\ &= E\{S_k|X\} \\ &= E\{S_k|X(\omega_0) X(\omega_1) \dots X(\omega_{L-1})\} \end{aligned} \quad (\text{II.41})$$

L'équation (II.41) représente la moyenne de la FDP a posteriori de \hat{S}_k . La FDP a postérieure $p(S_k|X)$ représente la FDP des amplitudes après observation des données (X). En revanche, $p(S_k)$ désigne la FDP a priori des amplitudes avant observation des données.

L'estimateur MMSE donné à l'équation (II.41), contrairement au filtre de Wiener, n'admet pas l'existence d'une relation linéaire entre les données observées et l'estimateur, mais il nécessite des connaissances sur les probabilités de distributions des coefficients de la TFD du signal de parole et du bruit. Supposons que nous connaissons a priori la distribution des coefficients TFD du signal de la parole et du bruit, nous pouvons évaluer $p(S_k|X)$ (moyenne de la FDP a posteriori de \hat{S}_k).

La mesure de la probabilité de distribution des coefficients TFD du signal de parole et cependant très difficile car le signal ou le bruit, ne sont pas des processus stationnaires et ergodiques. Ephraim et Malah [22] ont proposé un modèle statistique utilisant les propriétés statistiques asymptotiques de la Transformée de Fourier (TF), où il a été supposé que : (i) les coefficients de la TF (partie réelle et imaginaire) possèdent une distribution de probabilité gaussienne de moyenne nulle et de variance variable en raison de la non stationnarité du signal de parole, (ii) les coefficients de la TF sont statistiquement indépendants et donc non corrélés.

- Estimateur MMSE d'amplitude

Pour déterminer l'estimateur MMSE d'amplitude [22], premièrement nous calculons la FDP a posteriori $p(S_k|X(\omega_k))$ de S_k en utilisant la règle de Bayes suivante :

$$\begin{aligned} p(S_k|X(\omega_k)) &= \frac{p(X(\omega_k)|S_k) p(S_k)}{p(X(\omega_k))} \\ &= \frac{p(X(\omega_k)|S_k) p(S_k)}{\int_0^\infty p(X(\omega_k)|s_k)p(s_k)ds_k} \end{aligned} \quad (\text{II.42})$$

où s_k est une réalisation de la variable aléatoire S_k et $p(X(\omega_k))$ un facteur de normalisation.

Supposons que les coefficients de la TF sont statistiquement indépendants, tel que :

$$E\{S_k|X(\omega_0) X(\omega_1) \dots X(\omega_{N-1})\} = E\{S_k|X(\omega_k)\} \quad (\text{II.43})$$

En substituant (II.42) dans (II.41) on obtient l'estimateur de \hat{S}_k suivant :

$$\begin{aligned} \hat{S}_k &= E\{S_k|X(\omega_k)\} \\ &= \int_0^\infty s_k p(s_k|X(\omega_k))ds_k \\ &= \frac{\int_0^\infty s_k p(X(\omega_k)|s_k)p(s_k)ds_k}{\int_0^\infty p(X(\omega_k)|s_k)ds_k} \end{aligned} \quad (\text{II.44})$$

Sachant que

$$p(X(\omega_k)|S_k)p(S_k) = \int_0^{2\pi} p(X(\omega_k)|s_k, \theta_s)p(s_k, \theta_s)d\theta_s \quad (\text{II.45})$$

où θ_s est une réalisation de la variable aléatoire phase de $S(\omega_k)$, on obtient :

$$\hat{S}_k = \frac{\int_0^\infty \int_0^{2\pi} s_k p(X(\omega_k)|s_k, \theta_s) p(s_k, \theta_s) d\theta_s ds_k}{\int_0^\infty \int_0^{2\pi} p(X(\omega_k)|s_k, \theta_s) p(s_k, \theta_s) d\theta_s ds_k} \quad (\text{II.46})$$

$X(\omega_k)$ représente une somme de deux variables aléatoires complexes gaussiennes à moyennes nulles. Par conséquent, la FDP conditionnelle $p(X(\omega_k)|s_k, \theta_s)$ est aussi une gaussienne :

$$p(X(\omega_k)|s_k, \theta_s) = p_N(X(\omega_k) - S(\omega_k)) \quad (\text{II.47})$$

où $p_N(\cdot)$ est la FDP des coefficients de la TF du bruit $N(\omega_k)$.

L'équation (II.47) devient :

$$p(X(\omega_k)|s_k, \theta_s) = \frac{1}{\pi\sigma_n(k)} \exp\left\{-\frac{1}{\sigma_n(k)} |X(\omega_k) - S(\omega_k)|^2\right\} \quad (\text{II.48})$$

Pour des variables aléatoires gaussiennes complexes, nous savons que l'amplitude S_k et la phase $\theta_s(k)$ de la variable aléatoire $S(\omega_k)$ sont indépendantes, ce qui implique que $p(s_k, \theta_s) = p(s_k)p(\theta_s)$. La FDP de $\theta_s(k)$ est uniformément distribuée dans $(-\pi, \pi)$ et $p(s_k, \theta_s)$ peut être exprimée par :

$$p(s_k, \theta_s) = \frac{s_k}{\pi\sigma_s(k)} \exp\left\{-\frac{s_k^2}{\sigma_b(k)}\right\} \quad (\text{II.49})$$

En substituant les équations (II.48) et (II.49) dans (II.44), nous obtenons l'estimateur MMSE d'amplitude comme suit :

$$\hat{S}_k = \sqrt{\sigma_k} \Gamma(1.5) \Phi(-0.5, 1; -v_k) \quad (\text{II.50})$$

où $\Gamma(\cdot)$ est la fonction Gamma, $\Phi(a, b; c)$ est la fonction hypergéométrique confluyente (ou fonction de Kummer), σ_k est donnée par :

$$\sigma_k = \frac{\sigma_s(k)\sigma_n(k)}{\sigma_s(k) + \sigma_n(k)} = \frac{\sigma_s(k)}{1 + \xi_k} \quad (\text{II.51})$$

et v_k est définie par :

$$v_k = \frac{\xi_k}{1 + \xi_k} Y_k \quad (\text{II.52})$$

où Y_k et ξ_k sont définis par :

$$Y_k = \frac{X_k^2}{\sigma_b(k)} \quad (\text{II.53})$$

$$\xi_k = \frac{\sigma_s(k)}{\sigma_n(k)} \quad (\text{II.54})$$

Les termes ξ_k et Y_k , désignent respectivement les RSB a priori et a posteriori. ξ_k peut être considéré comme le vrai RSB de la $k^{\text{ième}}$ composante spectrale, par contre Y_k le RSB observé ou mesuré de la $k^{\text{ième}}$ composante spectrale. \hat{S}_k devient :

$$\hat{S}_k = \frac{\sqrt{v_k}}{Y_k} \Gamma(1.5) \Phi(-0.5, 1; -v_k) X_k \quad (\text{II.55})$$

En simplifiant les équations (II.52) à (II.54) et en exprimant $\Phi(-0.5, 1; -v_k)$ en termes de fonction de Bessel. L'estimateur MMSE devient :

$$\hat{S}_k = \frac{\sqrt{\pi} \sqrt{v_k}}{2 Y_k} \exp\left(-\frac{v_k}{2}\right) \left[(1 + v_k) I_0\left(\frac{v_k}{2}\right) + v_k I_1\left(\frac{v_k}{2}\right) \right] X_k \quad (\text{II.56})$$

Sachant que $\hat{S}_k = H(\xi_k, Y_k) X_k$, la fonction de suppression devient :

$$H_{MMSE}(\xi_k, Y_k) = \frac{\hat{S}_k}{X_k} = \frac{\sqrt{\pi} \sqrt{v_k}}{2 Y_k} \exp\left(-\frac{v_k}{2}\right) \left[(1 + v_k) I_0\left(\frac{v_k}{2}\right) + v_k I_1\left(\frac{v_k}{2}\right) \right] \quad (\text{II.57})$$

Nous remarquons que la fonction de suppression de l'estimateur MMSE d'amplitude est fonction des RSB a priori ξ_k et RSB a posteriori Y_k .

- Estimateur MMSE de la phase

Dans la section précédente, nous avons calculé un estimateur optimal, au sens MMSE, du spectre d'amplitude \hat{S}_k à partir de du signal bruité $X(\omega_k)$. Ce pendant pour pouvoir construire le spectre du signal rehaussé $\hat{S}(\omega_k) = \hat{S}_k \exp(j\hat{\theta}_s(k))$, il est nécessaire d'estimer la phase du signal. Pour cela, il a été montré [22] que l'estimateur optimal de la phase du signal est donné par :

$$\exp(j\hat{\theta}_s) = \exp(j\theta_x) \quad (\text{II.58})$$

- Estimation du RSB a priori

L'estimateur MMSE d'amplitude de l'équation (II.58) a été calculé en considérons que le RSB a priori ξ_k et la variance du bruit $\sigma_n(k)$ sont connus. Cependant en pratique, seul le signal bruité observé est connu. La variance du bruit peut être estimée par le biais du détecteur d'activité vocal (VAD : Voice Activity Detection), en supposant que le bruit est stationnaire. Ephraïm et Malah [22-23] ont proposé deux approches pour l'estimation du RSB à priori ξ_k : (a) une approche du maximum de vraisemblance (ML), (b) une approche de décision directe que nous présentons comme suit :

✓ Méthode du maximum de vraisemblance (ML)

Cette méthode peut être utilisée pour estimer le RSB a priori ξ_k en calculant, en premier lieu, le paramètre inconnu (supposé déterministe) $\sigma_s(k)$ et en estimant $\sigma_n(k)$ durant la période d'absence d'activité vocal. L'estimation de ξ_k est basée sur les L observations passées des amplitudes du signal observées $X_k(m) \triangleq \{X_k(m), X_k(m-1), \dots, X_k(m-L+1)\}$ obtenues à la $m^{\text{ième}}$ trame d'analyse et à l'indice de fréquence k . En supposant que les L observations sont statistiquement indépendantes et en utilisant le modèle gaussien, nous pouvons former la fonction de vraisemblance suivante :

$$p(\mathbf{X}_k(m); \sigma_s(k), \sigma_n(k)) = \prod_{j=0}^{L-1} \frac{1}{\pi(\sigma_s(k) + \sigma_n(k))} \exp \left\{ - \frac{X_k^2(m-j)}{\sigma_s(k) + \sigma_n(k)} \right\} \quad (\text{II.59})$$

En maximisant la fonction de vraisemblance par rapport à $\sigma_s(k)$, nous obtenons l'estimée de $\sigma_s(k)$ comme suit :

$$\begin{aligned} \hat{\sigma}_s(k, m) &= \begin{cases} \frac{1}{L} \sum_{j=0}^{L-1} X_k^2(m-j) - \sigma_n(k, m) & \text{si non négative} \\ 0 & \text{ailleurs} \end{cases} \\ &= \max \left(\frac{1}{L} \sum_{j=0}^{L-1} X_k^2(m-j) - \sigma_n(k, m), 0 \right) \end{aligned} \quad (\text{II.60})$$

En divisant les deux termes de l'équation (II.60) par $\sigma_n(k, m)$, nous obtenons :

$$\hat{\xi}_k(m) = \max \left(\frac{1}{L} \sum_{j=0}^{L-1} Y_k(m-j) - 1, 0 \right) \quad (\text{II.61})$$

où, $Y_k(m) \triangleq X_k^2(m) / \sigma_n(k, m)$ représente le RSB a posteriori de la trame m .

L'opérateur $\max(\cdot)$ est utilisé pour s'assurer que $\hat{\xi}_k(m)$ est aussi non négative. En pratique la moyenne glissante de l'équation (II.61) peut être remplacée par la relation récursive suivante :

$$\bar{Y}_k(m) = a\bar{Y}_k(m-1) + (1-a)\frac{Y_k(m)}{\beta} \quad (\text{II.62})$$

où $0 \leq a < 1$ est une constante de lissage, et $\beta \geq 1$ est un facteur de correction. L'estimateur ML final de ξ_k es donné par :

$$\hat{\xi}_k(m) = m \max(\bar{Y}_k(m) - 1, 0) \quad (\text{II.63})$$

✓ Méthode de Décision-Directe

La méthode Décision-Directe (DD) est basée sur la définition de ξ_k et sa relation avec le RSB a posteriori Y_k .

Sachant que :

$$\xi_k(m) = \frac{E\{S_k^2(m)\}}{\sigma_n(k, m)} \quad (\text{II.64})$$

On sait que ξ_k est relié à Y_k par :

$$\begin{aligned} \xi_k(m) &= \frac{E\{X_k^2(m) - N_k^2(m)\}}{\sigma_n(k, m)} \\ &= \frac{E\{X_k^2(m)\}}{\sigma_n(k, m)} - \frac{E\{N_k^2(m)\}}{\sigma_n(k, m)} \end{aligned} \quad (\text{II.65})$$

En combinant les équations (II.64) et (II.65), nous obtenons :

$$\xi_k(m) = E\left\{\frac{1}{2} \frac{S_k^2(m)}{\sigma_n(k, m)} + \frac{1}{2} [Y_k(m) - 1]\right\} \quad (\text{II.66})$$

L'estimateur final de ξ_k est calculé en exprimant récursivement l'équation (II.66) :

$$\hat{\xi}_k(m) = a \frac{\hat{S}_k^2(m-1)}{\sigma_n(k, m-1)} + (1-a)\max[Y_k(m) - 1, 0] \quad (\text{II.67})$$

où $0 < a < 1$ est un facteur de pondération et $\hat{S}_k^2(m-1)$ l'estimateur d'amplitude, obtenu dans les trames d'analyses précédentes. L'opérateur $\max(\cdot)$ est utilisé pour s'assurer que $\hat{\xi}_k(m)$ soit positive. Tel qu'il a été recommandé en [22], les conditions initiales du RSB a priori relatives à la première trame sont obtenues en mettant $m = 0$:

$$\hat{\xi}_k(0) = a + (1 - a)\max[Y_k(0) - 1, 0] \quad (\text{II.68})$$

avec $a = 0.98$.

Plusieurs travaux ont porté sur l'amélioration du RSB à priori $\hat{\xi}_k$ en agissant sur la formule finale de $\hat{\xi}_k$ dans l'approche de décision directe [67-69].

La méthode proposée par Ephraïm et Malah et une méthode de référence dans le domaine du rehaussement de la parole qui permet de réaliser de manière significative le compromis entre la réduction du bruit additif et la limitation du bruit musical.

Nous présentons dans la suite de ce chapitre, d'autres estimateurs qui ont été proposés dans la littérature, comme suit :

II.6.4. Estimateur LOG-MMSE

L'estimateur LOG-MMSE ou LSA [23], dérivé de l'estimateur MMSE, permet de minimiser l'erreur quadratique moyenne du spectre du logarithme d'amplitude comme suit :

$$e = E\left\{(\log S_k - \log \hat{S}_k)^2\right\} \quad (\text{II.69})$$

Cet estimateur peut être obtenu en calculant la moyenne conditionnelle du $\log S_k$, comme suit:

$$\log \hat{S}_k = E\{\log S_k | X(\omega_k)\} \quad (\text{II.70})$$

A partir de (II.70) on peut calculer \hat{S}_k comme suit :

$$\hat{S}_k = \exp(E\{\log S_k | X(\omega_k)\}) \quad (\text{II.71})$$

Il a été montré [23] que l'estimateur final LOG-MMSE est donné par :

$$\begin{aligned}\hat{S}_k &= \frac{\xi_k}{\xi_{k+1}} \exp \left\{ \frac{1}{2} \int_{v_k}^{\infty} \frac{e^{-t}}{t} dt \right\} X_k \\ &\triangleq H_{LSA}(\xi_k, v_k) X_k\end{aligned}\quad (\text{II.72})$$

H_{LSA} représente la fonction de suppression de l'estimateur LOG-MMSE.

II.6.5. Estimateur du Maximum a Posteriori (MAP)

Cet estimateur est basé sur la maximisation de la FDP a posteriori [70-71]. Bien que l'approche MMSE est axée sur le calcul de la moyenne de la FDP a posteriori $p(S_k|X(\omega_k))$ (c.à.d. $E\{S_k|X(\omega_k)\}$). L'estimateur MAP, par contre, vise à trouver le maximum de $p(S_k|X(\omega_k))$.

Dans l'approche MMSE, nous avons déterminé un estimateur du spectre d'amplitude et un estimateur de la phase, de manière indépendante et non conjointe. Avec le critère MAP, nous pouvons calculer les estimateurs du spectre d'amplitude et de la phase conjointement. Plus précisément, nous cherchons à trouver le maximum de la FDP a posteriori $p(s_k, \theta_k | X(\omega_k))$. l'estimateur MAP de l'amplitude et de la phase peut être calculé à partir de la relation suivante :

$$(\hat{s}_k, \hat{\theta}_s) = \arg \max_{s_k, \theta_k} p(s_k, \theta_s | X(\omega_k)) \quad (\text{II.73})$$

Il a été montré que l'estimateur MAP de la phase n'est autre que la phase du signal bruité, donnée comme suit:

$$\hat{\theta}_s = \theta_x \quad (\text{II.74})$$

Il a également été montré que l'estimateur du spectre d'amplitude basé sur le critère MAP est donné par :

$$\begin{aligned}\hat{S}_k &= \frac{\xi_k + \sqrt{\xi_k^2 + (1 + \xi_k) \frac{\xi_k}{Y_k}}}{2(1 + \xi_k)} X_k \\ &= H_{MAP}(\xi_k, Y_k) X_k\end{aligned}\quad (\text{II.75})$$

$H_{MAP}(\xi_k, Y_k)$ représente la fonction de suppression de l'estimateur MAP.

II.6.6. Estimateur MMSE d'ordre p

Dans cet estimateur [72], la fonction de suppression est calculée en fonction d'un facteur en puissance p . la valeur de $p=1$, nous permet d'obtenir l'estimateur MMSE d'origine du spectre d'amplitude de l'équation (II.56). Il a été remarqué que si p est grand, il résulte l'apparition d'un niveau de bruit résiduel assez élevé. En d'autres termes si p est petit, il en résulte une distorsion du signal. Pour cela, il a été procédé à l'adaptation du facteur p par rapport à chaque segment du signal.

II.6.7. Estimateurs MMSE basés sur des distributions non gaussiennes

Pour ces estimateurs [26-27] et [71], on suppose que les coefficients TFD du spectre d'amplitude du signal bruité et de la phase peuvent être modélisés par une loi Gamma ou de Laplace, ce qui a permis d'avoir des résultats améliorés par rapport à la méthode MMSE [22].

II.6.8. Estimateurs Bayésien perceptuels

Le concept de la fonction risque a permis à cet estimateur [73] d'incorporer les modèles psychoacoustiques dans l'estimation du spectre d'amplitude.

II.7. Méthodes de rehaussement incorporant la probabilité d'absence de la parole

Dans les méthodes précédentes, il a été supposé que le signal de parole est continu sur toute la durée d'observation. Cependant, en pratique, le signal de parole contient beaucoup de pauses, même durant l'activité vocale. Il s'ensuit qu'une règle de suppression du bruit plus efficace peut être obtenue si nous supposons que la parole peut être représentée par un modèle à deux états qui concerne la présence ou l'absence du signal de parole durant la durée d'observation. Dans cette partie, nous présentons quelques méthodes qui supposent que le signal de parole peut ne pas être présent durant toute la durée d'observation. Intuitivement, cela revient à multiplier la règle de suppression du bruit par un terme qui permet d'estimer la probabilité de présence de la parole dans une bande de fréquence donnée.

Le modèle à deux états de la parole peut être exprimé en utilisant le modèle des hypothèses binaires [21] :

$$\begin{aligned}
 H_0^k &: \text{absence de la parole} : |X(\omega_k)| = |N(\omega_k)| \\
 H_1^k &: \text{présence de la parole} : |X(\omega_k)| = |S(\omega_k) + N(\omega_k)| \\
 &= |S_k e^{j\theta} + N(\omega_k)|
 \end{aligned} \tag{II.76}$$

où H_0^k représente l'hypothèse nulle d'absence du signal de parole dans la bande de fréquence k , et H_1^k représente l'hypothèse non nulle de présence de la parole. En vue d'incorporer le modèle binaire précédent dans l'estimateur Bayésien, nous utilisons une moyenne pondérée des deux estimateurs à savoir : la moyenne pondérée par la probabilité de présence du signal de parole et la moyenne pondérée par la probabilité d'absence du signal de parole. Ainsi, si l'estimateur original MMSE d'amplitude S_k a pour forme $\hat{S}_k = E(S_k|X_k)$, le nouveau estimateur devient :

$$\hat{S}_k = E(S_k|X_k, H_1^k)P(H_1^k|X_k) + E(S_k|X_k, H_0^k)P(H_0^k|X_k) \quad (\text{II.77})$$

où $P(H_1^k|X_k)$ désigne la probabilité conditionnelle de présence de la parole, dans la bande de fréquence k , sachant X_k . De même, $P(H_0^k|X_k)$ représente la probabilité conditionnelle d'absence de la parole, dans la bande de fréquence k sachant X_k . Le terme $E(S_k|X_k, H_0^k)$ dans (II.77) est nulle car il représente la valeur moyenne de S_k sachant X_k et H_0^k (absence du signal de parole). En conséquence, l'estimateur MMSE de (II.77) devient :

$$\hat{S}_k = E(S_k|X_k, H_1^k)P(H_1^k|X_k) \quad (\text{II.78})$$

$E(S_k|X_k, H_1^k)$ peut être approximé par l'estimateur ML de (II.29). L'équation (II.78) devient donc :

$$\hat{S}_k = \frac{1}{2} \left[X_k + \sqrt{X_k^2 - N_k^2} \right] P(H_1^k|X_k) \quad (\text{II.79})$$

Le calcul de $P(H_1^k|X_k)$ peut être effectué par l'utilisation de la règle de Bayes :

$$\begin{aligned} P(H_1^k|X_k) &= \frac{p(X_k|H_1^k)P(H_1)}{p(X_k)} \\ &= \frac{p(X_k|H_1^k)P(H_1)}{p(X_k|H_1^k)P(H_1) + p(X_k|H_0^k)P(H_0)} \end{aligned} \quad (\text{II.80})$$

Sous l'hypothèse H_0 , $X_k = |N(\omega_k)| = N_k$ est un bruit gaussien avec moyenne nulle et variance $\sigma_n(k)$. Sous ces conditions, $p(X_k|H_0^k)$ peut avoir une distribution de probabilité de Rayleigh donnée par :

$$p(X_k|H_0^k) = \frac{2X_k}{\sigma_n(k)} \exp\left(-\frac{X_k^2}{\sigma_n(k)}\right) \quad (\text{II.81})$$

Sous l'hypothèse H_1 , $X_k = |S(\omega_k) + N(\omega_k)|$ et $p(X_k|H_1^k)$ suit la loi de Rice donnée par :

$$p(X_k|H_1^k) = \frac{2X_k}{\sigma_n(k)} \exp\left(-\frac{X_k^2 + S_k^2}{\sigma_n(k)}\right) I_0\left(\frac{2X_k S_k}{\sigma_n(k)}\right) \quad (\text{II.82})$$

Où $I_0(\cdot)$ est la fonction modifiée de Bessel du premier ordre. En substituant les équations (II.81) et (II.82) dans l'équation (II.80) et en supposant que $P(H_1) = P(H_0) = \frac{1}{2}$, nous obtenons :

$$p(H_1^k|X_k) = \frac{\exp(-\xi_k) I_0(2\sqrt{\xi_k Y_k})}{1 + \exp(-\xi_k) I_0(2\sqrt{\xi_k Y_k})} \quad (\text{II.83})$$

où $\xi_k = S_k^2/\sigma_n(k)$ et $Y_k = X_k^2/\sigma_n(k)$.

La figure II. 6 présente le tracé de (II.83) en fonction du RSB a posteriori Y_k , où on peut remarquer que lorsque Y_k est grand, $p(H_1^k|X_k)$ est égale à 1 ce qui signifie que la parole est vraisemblablement présente. Lorsque Y_k est petit, la valeur de $p(H_1^k|X_k)$ dépend de la valeur de ξ_k . Si ξ_k est grand et Y_k est petit, alors $p(H_1^k|X_k)$ tend vers zéro.

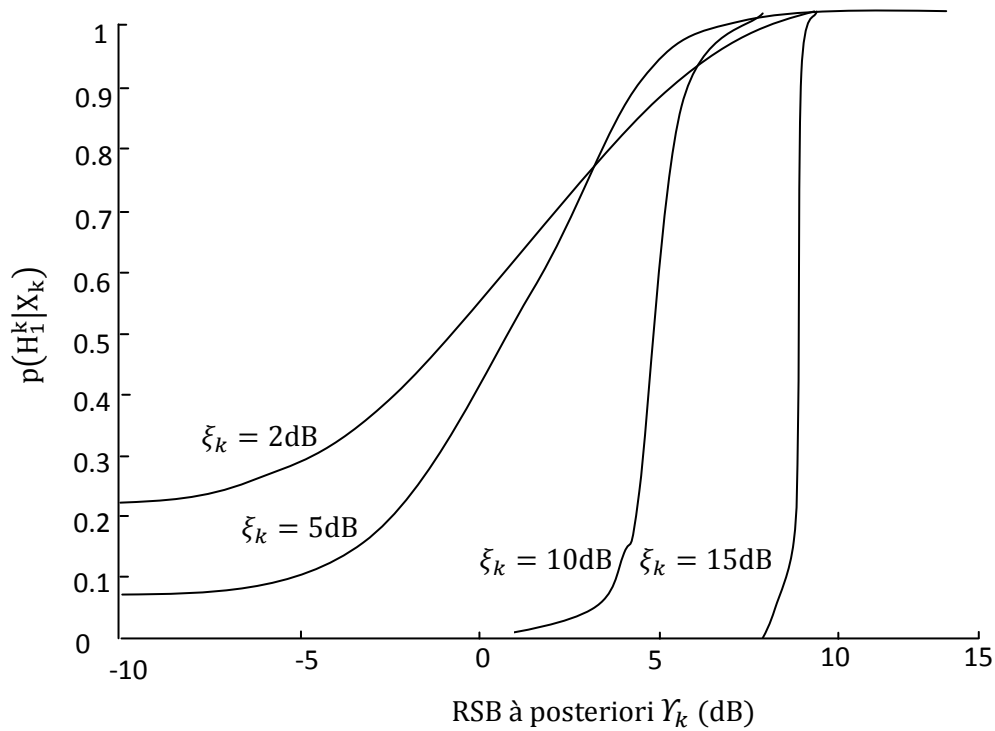


Figure II. 6 : Tracer de la probabilité a posteriori de présence de la parole $p(H_1^k|X_k)$, en fonction du RSB a posteriori Y_k [1]

II.8. Rehaussement de la parole par filtrage de Kalman

Le filtrage de Kalman [101] est une technique de rehaussement qui consiste à améliorer la qualité et l'intelligibilité du signal de la parole, modélisé par un processus autorégressif (AR).

Ce filtre est caractérisé par :

- Modélisation autorégressive du signal de la parole ;
- Inexistence du phénomène de bruit musical ;
- Technique adaptée aux signaux de parole non-stationnaire ;
- Le filtrage de Kalman peut être utilisé sur des bruits blancs ou colorés.

On suppose tout d'abord que le signal de la parole $s(n)$ est modélisé par un processus autorégressif (AR) d'ordre p :

$$s(n) = \sum_{i=1}^p a_i s(n-i) + u(n) \quad (\text{II.84})$$

Le signal bruité de (II.1) peut être exprimé par une forme discrète comme suit :

$$x(n) = s(n) + n(n) \quad (\text{II.85})$$

avec :

a_i : Coefficients de la prédiction linéaire (LPC).

$u(n)$: Bruit de processus.

$n(n)$: Bruit blanc additif.

On peut représenter le système dans l'espace d'état comme suit :

$$\begin{cases} y(n) = Fy(n-1) + Gu(n) \\ x(n) = Hy(n) + n(n) \end{cases} \quad (\text{II.86})$$

avec : $y(n)$: Vecteur d'état ($p \times 1$), constitué par les p dernières valeurs du signal $s(n)$ est donné par :

$$y(n) = [s(n-p+1) \dots s(n)]^T \quad (\text{II.87})$$

F est une matrice de transition ($p \times p$) donnée par :

$$F = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \\ a_p & a_{p-1} & a_{p-2} & \cdots & a_1 \end{bmatrix} \quad (\text{II.88})$$

G est un vecteur d'entrée ($p \times 1$) donné par :

$$G = [0, \dots, 0, 1]^T \quad (\text{II.89})$$

H est un vecteur d'observation ($1 \times p$) donné par :

$$H = [0, \dots, 0, 1] = G^T \quad (\text{II.90})$$

$u(n)$ et $n(n)$ sont des bruits blancs gaussiens indépendants de moyennes nulles et de matrices de covariance Q et R respectivement.

Le filtre de Kalman permet de calculer \hat{y}_n qui est l'estimation du vecteur d'état y_n en utilisant les équations récursives suivantes :

$$\hat{y}_n^- = F \cdot \hat{y}_{n-1}^- \quad (\text{II.91})$$

$$P_n^- = F \cdot P_{n-1}^- \cdot F^T + G \cdot Q \cdot G^T \quad (\text{II.92})$$

$$K_n = P_n^- \cdot H [H \cdot P_n^- \cdot H^T + R_n]^{-1} \quad (\text{II.93})$$

$$\hat{y}_n = \hat{y}_n^- + K_n [x_n - H \cdot \hat{y}_n^-] \quad (\text{II.94})$$

$$P_n = [I - K_n \cdot H] \cdot P_n^- \quad (\text{II.95})$$

avec :

\hat{y}_n^- : Estimation du vecteur d'état y_n , au sens MMSE, donnée par les observations précédentes $x(1), \dots, x(n-1)$.

P_n^- : Matrice de covariance de l'erreur de l'état prédit.

K_n : Gain de Kalman.

\hat{y}_n : Estimation filtrée du vecteur d'état y_n .

P_n : Matrice de covariance de l'erreur de l'état filtré.

L'estimation de la parole propre peut être obtenue à partir de l'estimation filtrée du vecteur d'état par la relation :

$$\hat{s}(n) = H\hat{y}_n \quad (\text{II.96})$$

Bien que la matrice de transition soit $H = [0, \dots, 0, 1]$, on peut considérer que la dernière composante du vecteur d'état estimé représente l'estimation filtrée de Kalman du signal de la parole $s(n)$.

Plusieurs méthodes d'estimations peuvent être utilisées pour estimer les coefficients AR $\{a_1, \dots, a_p\}$ et les variances des bruits Q et R pour chaque segment stationnaire de la parole.

Le concept de Kalman a été également étalé au cas du bruit coloré [102], où il a été question dans ce cas de la nécessité de modéliser, à la fois le signal de la parole $s(n)$ et le bruit additif $n(n)$ par deux modèles autorégressifs (AR) d'ordre p et q respectivement.

Le rehaussement de la parole par filtrage de Kalman, nécessite l'estimation des paramètres suivants : LPC et matrice de covariance. Dans ce cas, la qualité du signal rehaussé est étroitement liée à la fiabilité de l'estimation de ces paramètres. Cette estimation est effectuée, dans le cas idéal, à partir de la parole propre. Cependant en pratique, elle demeure impossible du fait que la parole propre est à priori inconnue et seule la parole bruitée est disponible pour l'extraction des différents paramètres.

Une estimation imprécise des paramètres dégrade les performances du filtre de Kalman. Pour cela, on utilise l'algorithme EM [103] qui sert à estimer et corriger récursivement ces paramètres.

II.9. Conclusion

Ce chapitre a été consacré à l'étude des méthodes de rehaussement de la parole de l'état de l'art, notamment les méthodes SS et ceux basés sur la théorie de l'estimation statistique. Nous avons décrit également le phénomène du bruit musical, qui est une caractéristique des méthodes SS et nous avons également étudié les différentes règles de suppression ainsi que leur robustesse quant à l'élimination du bruit musical.

Nous nous sommes également intéressés aux algorithmes qui prennent en charge le concept des hypothèses binaires avant de clore ce chapitre par l'étude du rehaussement de la parole par filtrage de Kalman.

Dans le prochain chapitre, nous détaillons les méthodes de rehaussement basées sur la décomposition en sous-espaces, le principe de l'Analyse en Composante Principales (PCA : Principal Component Analysis) ainsi que le critère de la Variance de l'Erreur de Reconstruction (VRE : Variance of the Reconstruction Error) qui est destiné pour la sélection du nombre optimal des Composantes Principales à retenir pour la reconstruction du signal propre.

Chapitre III

Rehaussement de la parole par décomposition en sous-espaces du signal

III.1. Introduction

Les algorithmes de rehaussement de la parole présentés dans le chapitre précédent étaient basés sur le traitement du signal et de l'estimation statistique. Dans ce chapitre, nous allons aborder d'autres méthodes fondées sur la théorie de l'algèbre linéaire. Le principe de ces algorithmes réside sur le fait que le signal propre pourrait se limiter à un seul sous-espace issu de l'espace euclidien bruité. Par conséquent, connaissant la méthode de décomposition de l'espace vectoriel bruité en un sous-espace occupé par le signal propre et un autre sous-espace occupé par le bruit on peut estimer le signal propre par l'annulation de la composante du vecteur bruité occupant le sous-espace bruit. La décomposition peut être effectuée en utilisant des techniques de factorisation des matrices orthogonales, en particulier la décomposition en valeurs singulières (SVD : Singular Value Decomposition) [53,74-75] qui permet d'obtenir des bases orthonormées de quatre sous-espaces fondamentaux : l'espace des colonnes de la matrice (sous-espace du signal), l'espace nul gauche, l'espace des lignes et l'espace nul. Les matrices orthogonales obtenues à partir de la SVD ou également par décomposition en valeurs propres (EVD) [76] peuvent être considérées comme des transformations destinées à être appliquées à la matrice d'observations (signal bruité). En effet, La matrice formée par les vecteurs propres, après décomposition en valeurs singulières de la matrice de covariance du signal (ou d'autocorrélation) est souvent appelée Transformée de Karhunen-Loève (KLT : Karhunen-Loève Transform).

La méthode de décomposition en sous-espaces a été initialement proposée par Tufts et al. [77] où il a été utilisé la SVD de la matrice des observations pour l'estimation du signal propre. Plus tard, Cadzow [78] a présenté une structure plus générale en supposant que le signal propre présente quelques propriétés bien définies obéissant à un certain modèle. Le rehaussement de la parole est donc obtenu en mappant le signal bruité sur l'espace des

signaux possédant la même structure que le signal propre. Cette théorie est considérée comme une référence pour toutes les méthodes de rehaussement de la parole en sous-espace.

La première étape dans la réduction du bruit consiste à annuler le sous-espace du bruit [40]. Cependant, l'amélioration de cette méthode nécessite le contrôle ou la suppression de la contribution du bruit dans le sous-espace (signal + bruit), ce qui est effectué par d'autres estimateurs que nous détaillons dans la suite de ce chapitre.

L'estimateur de la variance minimale (MV : Minimum Variance) [79] permet une meilleure estimation linéaire des données propres en connaissant le rang de la matrice du signal propre et la variance du bruit blanc. Plus tard, il a été remarqué dans [32] que la distorsion du signal et le bruit résiduel ne peuvent pas être minimisés simultanément. Pour cela, deux nouveaux estimateurs linéaires ont été proposés à savoir : (i) Contrainte du Domaine Temporel (TDC : Time Domain Constrained) et (ii) Contrainte du Domaine Spectral (SDC : Spectral Domain Constrained) ayant pour objectif de maintenir le niveau du bruit résiduel inférieur à un seuil choisi, tout en minimisant la distorsion du signal.

Bien que les méthodes en sous-espaces ont concerné la réduction des bruits blancs, ces dernières ont été étendues aux bruits colorés à condition que la matrice de covariance du bruit soit connue (estimée) [80].

L'effet de masquage du bruit [41] a été également utilisé dans le développement de nouveaux algorithmes de rehaussement de la parole en sous-espaces qui consistent à réduire l'effet perceptuel du bruit musical. Ce modèle est basé sur le fait que le bruit additif est inaudible pour l'oreille humaine tant qu'il est inférieur à un certain seuil de masquage.

III.2. Formalismes mathématiques

III.2.1. Définitions

Le signal de parole $s(t)$ est une variable stochastique qui peut être considérée comme une réalisation d'un processus stochastique $\{S(t)\}$ supposé non-stationnaire qui peut être considéré, dans de courtes durées de 10 à 30ms, comme stationnaire au sens large. Nous considérons que les processus stochastiques stationnaires au sens large (WSS) $\{X(t)\}$, $\{N(t)\}$, et $\{S(t)\}$ qui génèrent respectivement les variables $x(t)$, $n(t)$ et $s(t)$ sont des processus ergodiques au sens large (WSE).

Soit $x(t) \in \mathcal{R}^K$ le signal d'observation (bruité) composé par le signal propre $s(t) \in \mathcal{R}^K$ et par le bruit $n(t) \in \mathcal{R}^K$, comme suit:

$$x(t) = s(t) + n(t) \quad (\text{III.1})$$

où
$$x(t) = [x_1, x_2, \dots, x_K]^T \quad (\text{III.2})$$

Nous arrangeons le vecteur bruité $x(t)$ ($K \times 1$) dans une matrice d'observation de Hankel de dimension $M \times N$, $X_{M \times N}(t)$, où $K = M + N - 1$, comme suit :

$$X_{M \times N}(t) = \begin{pmatrix} x_t & x_{t-1} & \dots & x_{t-N+1} \\ x_{t-1} & x_{t-2} & \dots & x_{t-N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{t-M+1} & x_{t-M} & \dots & x_{t-M-N+2} \end{pmatrix} \quad (\text{III.3})$$

III.2.2. Matrice de covariance

Nous obtenons à partir de $x(t)$ la matrice de covariance R_{xx} définie comme suit:

$$R_{xx} = E\{xx^T\} \quad (\text{III.4})$$

avec R_{xx} :

- Symétrique de valeurs réelles : $R_{xx} = R_{xx}^T$;
- Positive : $\Psi^T R_{xx} \Psi \geq 0, \forall \Psi$.

Nous pouvons calculer la valeur estimée de la matrice de covariance comme suit [66] :

$$\hat{R}_{xx} = \begin{cases} \frac{1}{N-1} X^T X \in \mathcal{R}^{M \times M} \\ \frac{1}{M-1} X^T X \in \mathcal{R}^{N \times N} \end{cases} \quad (\text{III.5})$$

\hat{R}_{xx} converge vers la matrice de corrélation (principe d'ergodicité) [66] :

$$R_{xx} = \lim_{M \rightarrow \infty} \frac{1}{M-1} X^T X \in \mathcal{R}^{N \times N} \quad (\text{III.6})$$

Ainsi, pour des matrices d'observations $X \in \mathcal{R}^{M \times N}$, où $M > N$, on choisit $\frac{1}{M-1} X^T X \in \mathcal{R}^{N \times N}$ comme étant la matrice de faible dimension qui constitue la meilleure estimation de la matrice de covariance.

III.2.3. Valeurs propres et vecteurs propres

Une valeur propre ou valeur caractéristique est définie comme étant un scalaire justifiant la relation suivante :

$$X\mathbf{v} = \lambda\mathbf{v} \quad (\text{III.7})$$

Le vecteur $\mathbf{v} \neq 0$ qui satisfait la relation (III.7) est appelé vecteur propre de X associé avec la valeur propre λ . L'équation (III.7) peut être représentée comme suit :

$$(X - I\lambda)\mathbf{v} = 0 \quad (\text{III.8})$$

où I est la matrice identité.

A partir de (III.7), on déduit la relation suivante :

$$R_{xx}q = \lambda q \quad (\text{III.9})$$

où q_1, q_2, \dots, q_N et $\lambda_1, \lambda_2, \dots, \lambda_N$ représentent respectivement les vecteurs propres et valeurs propres de R_{xx} .

Nous savons que lorsqu'une matrice carrée $N \times N$ est symétrique, les valeurs propres associées λ_i , $i = 1, \dots, N$ sont réelles et les vecteurs propres q_i , $i = 1, \dots, N$ forment des bases orthonormées :

$$\begin{aligned} \lambda_i &\in \mathcal{R}, \quad i = 1, \dots, N \\ q_i^T q_j &= \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \end{aligned} \quad (\text{III.10})$$

III.3. Rehaussement de la parole par décomposition en valeurs propres (EVD)

Les valeurs et vecteurs propres définis dans l'équation (III.9) peuvent être utilisés dans le cadre de la décomposition en valeurs propres de la matrice d'autocorrélation R_{xx} comme suit [32] :

$$\begin{cases} R_{xx} = Q\Lambda_x Q^T \\ \Lambda_x \triangleq \text{diag}\{\lambda_x(1), \dots, \lambda_x(N)\} \end{cases} \quad (\text{III.11})$$

Dans la plupart des techniques de rehaussement en sous-espaces, on considère que le bruit est blanc de variance σ_n^2 .

On peut définir la matrice Q comme :

$$Q = [q_1 q_2 \dots q_N] \in \mathcal{R}^{N \times N} \quad (\text{III.12})$$

Les vecteurs propres sont orthogonaux en raison de la symétrie dans R_{xx} et les valeurs propres sont arrangées dans la matrice diagonale par ordre décroissant comme suit:

$$\lambda_x(k) = \begin{cases} \lambda_s(k) + \sigma_n^2, & \text{si } k = 1, \dots, N \\ \sigma_n^2 & \text{si } k = N + 1, \dots, K \end{cases} \quad (\text{III.13})$$

Sachant que :

- Pour une matrice semi-définie positive : $\lambda_x(1) \geq \lambda_x(2) \geq \dots \geq \lambda_x(N) \geq 0$;
- Pour une matrice définie positive : $\lambda_x(1) \geq \lambda_x(2) \geq \dots \geq \lambda_x(N) > 0$.

Aussi, la décomposition en valeurs propres de R_{ss} [32] est donnée comme suit :

$$\begin{cases} R_{ss} = Q \Lambda_s Q^T \\ \Lambda_s \triangleq \text{diag}\{\lambda_s(1), \dots, \lambda_s(N)\} \end{cases} \quad (\text{III.14})$$

Les valeurs propres de Λ_x ainsi que les vecteurs propres associés sont considérés respectivement comme les valeurs et vecteurs propres principaux de R_{xx} .

La décomposition en valeurs propres de la matrice d'autocorrélation du bruit peut être exprimée par [32]:

$$\begin{cases} R_{nn} = Q \Lambda_n Q^T \\ \Lambda_n = \sigma_n^2 I, \text{ avec } I \text{ matrice identité} \end{cases} \quad (\text{III.15})$$

A partir de (III.11), (III.14) et (III.1) on peut remarquer que : $\Lambda_s = \Lambda_x - \sigma_n^2 I$

On peut également exprimer R_{xx} comme:

$$R_{xx} = R_{ss} + R_{nn} , \quad (\text{III.16})$$

En remplaçant (III.14) et (III.15) dans (III.16), on obtient :

$$R_{xx} = Q(\Lambda_s + \sigma_n^2 I)Q^T \quad (\text{III.17})$$

Notons que $Q = [Q_1, Q_2]$, où Q_1 est la matrice ($K \times N$) des vecteurs propres principaux de R_{xx} :

$$Q_1 = \{q_k : \lambda_x(k) > \sigma_n^2\} \quad (\text{III.18})$$

Sachant que Q est orthonormée, on aura donc la relation suivante :

$$I = Q_1 Q_1^T + Q_2 Q_2^T \quad (\text{III.19})$$

Ainsi, on peut décomposer le vecteur x (signal bruité) comme suit :

$$x = Q_1 Q_1^T x + Q_2 Q_2^T x \quad (\text{III.20})$$

où :

- $Q_1 Q_1^T$ représente la projection de x sur l'espace signal;
- $Q_2 Q_2^T$ représente la projection de x sur l'espace bruit.

Les coefficients des vecteurs des deux projections $Q_1^T x$ et $Q_2^T x$ sont obtenus à partir de la Transformée de Karhunen-Loève de x ($Q^T x$), avec :

$$\begin{aligned} E\{Q^T x\} &= 0 \\ cov(Q^T x) &= diag[\Lambda_s + \sigma_n^2 I] \end{aligned} \quad (\text{III.21})$$

III.4. Estimateur linéaire du signal

Nous présentons, dans cette section, deux estimateurs linéaires qui permettent de minimiser la distorsion du signal tout en limitant l'énergie et le spectre du bruit résiduel. Nous présentons les estimateurs TDC et SDC qui consistent à maintenir l'énergie du bruit résiduel en dessous d'un seuil donné moyennant deux contraintes temporelle et spectrale.

III.4. 1. Estimateur TDC

Par application d'un filtre linéaire $H(K \times K)$ sur le signal bruité x , on peut obtenir la valeur estimée \hat{s} du signal propre s comme suit [32]:

$$\hat{s} = Hx = Hs + Hn \quad (\text{III.22})$$

l'erreur résiduelle est définie comme suit :

$$r = \hat{s} - s = (H - I)s + Hn = r_s + r_n \quad (\text{III.23})$$

$r_s \triangleq (H - I)s$ et $r_n \triangleq Hn$ représentent respectivement la distorsion du signal et du bruit résiduel.

Le rehaussement du signal corrompu par un bruit additif est réalisé en considérant que le signal propre est concentré dans un sous-espace de dimension $r < N$ (sous-espace signal), tandis que le bruit occupe le sous-espace de dimension $N - r$ restant. Par conséquent, l'opération de réduction du bruit est effectuée en annulant le sous-espace du bruit et en minimisant la contribution du bruit dans le sous-espace bruit. Cette opération, nécessite a priori des connaissances sur la dimension du signal pour le définir correctement (ordre optimal de la reconstruction).

Nous définissons l'énergie des vecteurs r_s et r_n respectivement comme:

$$\epsilon_s^2 = \text{tr}E\{r_s r_s^T\} = \text{tr}\{(H - I)R_s(H - I)^T\} \quad (\text{III.24})$$

$$\epsilon_n^2 = \text{tr}E\{r_n r_n^T\} = \sigma_n^2 \text{tr}\{HH^T\} \quad (\text{III.25})$$

où (tr) désigne la trace de la matrice.

L'estimateur linéaire de r_n est obtenu par:

$$\begin{aligned} & \min_H \epsilon_s^2 \\ & \text{qui respecte : } \frac{1}{K} \epsilon_n^2 \leq \alpha \sigma_n^2, \quad 0 \leq \alpha \leq 1 \end{aligned} \quad (\text{III.26})$$

L'équation (III.26) signifie que l'estimateur dérivé qui permet de minimiser la distorsion du signal et soumis à une contrainte relative au niveau admissible du bruit résiduel $\alpha \sigma_n^2$.

L'estimateur optimal H cherché satisfait l'équation du gradient du lagrangien [32]:

$$L(H, \mu_L) = \epsilon_s^2 + \mu(\epsilon_n^2 - \alpha K \sigma_n^2) \quad (\text{III.27})$$

et:
$$\mu(\epsilon_n^2 - \alpha K \sigma_n^2) = 0 \text{ pour } \mu \geq 0 \quad (\text{III.28})$$

la résolution de l'équation $\nabla L(H, \mu_L) = 0$ conduit à l'estimateur optimal (TDC) suivant [32]:

$$H_{TDC} = R_{SS}(R_{SS} + \mu \sigma_n^2 I)^{-1} \quad (\text{III.29})$$

où μ est le multiplicateur de Lagrange.

à partir de l'équation (III.28), on peut déduire que :

$$\epsilon_n^2 = \alpha K \sigma_n^2 \quad (\text{III.30})$$

En substituant (III.29) dans (III.30) on trouve que μ satisfait la relation suivante :

$$\alpha = \frac{1}{K} \text{tr}\{R_{SS}^2 (R_{SS} + \mu \sigma_n^2 I)^{-2}\} \quad (\text{III.31})$$

Le filtre optimal H_{TDC} peut être exprimé en appliquant la décomposition en valeurs propres de l'équation (III.14) à (III.30) comme suit [32]:

$$H_{TDC} = Q \begin{bmatrix} G_\mu & 0 \\ 0 & 0 \end{bmatrix} Q^T \quad (\text{III.32})$$

où
$$G_\mu = \Lambda_s (\Lambda_s + \mu \sigma_n^2 I)^{-1} \quad (\text{III.33})$$

L'estimée du signal devient :

$$\hat{S}_{TDC} = H_{TDC} X \quad (\text{III.34})$$

\hat{S}_{TDC} est obtenu par application de la transformée KLT au signal bruité et modification de $Q^T X$ par une fonction gain et par la KLT inverse (IKLT). Ces modifications du gain permettent d'annuler les composantes bruitées confinées dans le sous-espace du bruit.

Le filtre optimal devient :

$$H_{TDC} = Q_1 G_\mu Q_1^T \quad (\text{III.35})$$

On déduit, à partir de (III.35), le signal estimé comme suit :

$$\hat{S} = Q_1 G_\mu Q_1^T X \quad (\text{III.36})$$

L'estimateur linéaire (III.35) peut être également exprimé de manière explicite comme suit:

$$H_{TDC} = \sum_{k=1}^N g_\mu(k) q_k q_k^T \quad (\text{III.37})$$

Où $g_\mu(k)$ désigne le $k^{\text{ième}}$ élément diagonal de G_μ donné par :

$$g_\mu(k) = \frac{\lambda_s(k)}{\lambda_s(k) + \mu \sigma_n^2} \quad (\text{III.38})$$

La figure III.1. représente le schéma bloc de l'estimateur linéaire :

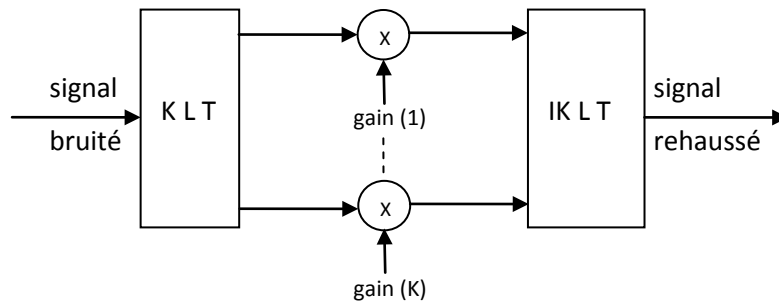


Figure III.1 : Estimateur linéaire de la méthode de décomposition en sous-espace [1]

III.4.2. Estimateur SDC

Nous dérivons un autre estimateur [32] linéaire qui permet la minimisation de la distorsion du signal et qui est supposé soumis à des contraintes sur le spectre du bruit résiduel (considéré comme identique au spectre de la parole), où on peut considérer que le bruit résiduel peut être masqué par le signal de la parole. La $k^{\text{ième}}$ composante spectrale du bruit résiduel est donnée par $q_k^T r_n$.

Aussi, on considère les deux intervalles suivants :

- Pour $k = 1, \dots, N$, on a $q_k^T r_n \leq \alpha_k \sigma_n^2$, avec $0 < \alpha_k \leq 1$;
- Pour $k = N + 1, \dots, K$, on a $q_k^T r_n = 0$.

Sachant que l'énergie du signal dans le sous-espace du bruit est nulle, le filtre H devient comme suit :

$$\begin{aligned} & \min_H \epsilon_s^2 \\ & E \{ |q_k^T r_n|^2 \} \leq \alpha_k \sigma_n^2, \quad k = 1, \dots, N \\ \text{qui respecte :} & \\ & E \{ |q_k^T r_n|^2 \} = 0, \quad k = N + 1, \dots, K \end{aligned} \quad (\text{III.39})$$

En outre, le filtre optimal H_{SDC} doit également satisfaire l'équation du gradient suivante:

$$HR_{SS} + \sigma_n^2 LH - R_{SS} = 0 \quad (\text{III.40})$$

$$\text{où} \quad L = Q \Lambda_\mu Q^T \quad (\text{III.41})$$

$\Lambda_\mu = \text{diag}(\mu_{L1}, \dots, \mu_{LK})$ représente la matrice diagonale du multiplicateur de Lagrange.

En substituant (III.14) dans (III.40), on obtient :

$$(I - \mathbb{U})A_s - \sigma_n^2 \Lambda_\mu = 0 \quad (\text{III.42})$$

avec $\mathbb{U} \triangleq Q^T H Q$.

En supposant que \mathbb{U} est diagonale, une solution possible pour l'équation (III.42) peut être obtenue par :

$$\mathbb{U}_{kk} = \begin{cases} \frac{\lambda_s(k)}{\lambda_s(k) + \mu \sigma_n^2}, & k = 1, \dots, N \\ 0, & k = N + 1, \dots, K \end{cases} \quad (\text{III.43})$$

Pour cette valeur de \mathbb{U} , nous avons :

$$E \{ |q_k^T r_n|^2 \} = \begin{cases} \sigma_n^2 \mathbb{U}_{kk}^2, & k = 1, \dots, N \\ 0, & k = N + 1, \dots, K \end{cases} \quad (\text{III.44})$$

Si la première contrainte de l'équation (III.39) est satisfaite par $(E \{ |q_k^T r_n|^2 \} = \alpha_k \sigma_n^2)$, on aura donc :

$$\sigma_n^2 \mathbb{U}_{kk}^2 = \alpha_k \sigma_n^2 \Rightarrow \mathbb{U}_{kk} = (\alpha_k)^{1/2}, \quad k = 1, \dots, N \quad (\text{III.45})$$

et

$$\mu_{Lk} = \frac{\lambda_s(k)}{\sigma_n^2} [(\alpha_k)^{1/2} - 1], \quad k = 1, \dots, N \quad (\text{III.46})$$

A partir de (III.43) et (III.45), le filtre optimal H_{SDC} devient :

$$\begin{aligned} H_{SDC} &= Q\mathbb{U}Q^T \\ \mathbb{U} &= \text{diag}(q_{11}, \dots, q_{KK}) \\ \mathbb{U}_{kk} &= \begin{cases} \alpha_k^{1/2}, & k = 1, \dots, N \\ 0, & k = N + 1, \dots, K \end{cases} \end{aligned} \quad (\text{III.47})$$

On peut remarquer de l'équation (III.47) que le choix de $\{\alpha_k\}$ définit complètement les gains de l'estimateur. Nous supposons dans cette approche que le bruit d'entrée est blanc de Densité Spectrale de Puissance (DSP) σ_n^2 et que le spectre non nul du bruit résiduel de sortie est $\alpha_k \sigma_n^2$. Il a été démontré [40] que théoriquement, les $\{\alpha_k\}$ peuvent être choisis indépendamment de la statistique du signal et du bruit comme suit :

$$\alpha_k = \left(\frac{\lambda_s(k)}{\lambda_s(k) + \sigma_n^2} \right)^\gamma \quad (\text{III.48})$$

où $\gamma \geq 1$ est une valeur constante déterminée expérimentalement destinée à effectuer le compromis entre la suppression du bruit et la distorsion du signal estimé au sens où si γ augmente, le niveau du bruit résiduel autorisé diminue et le niveau de distorsion du signal augmente. α_k peut être exprimée sous une autre forme comme suit :

$$\alpha_k = \exp\{-\nu \sigma_n^2 / \lambda_s(k)\} \quad (\text{III.49})$$

ν est un paramètre permettant le compromis entre la suppression du bruit et la distorsion du signal, identique à γ .

III.5. Rehaussement de la parole par Décomposition en Valeurs Singulières (SVD)

La décomposition en valeurs singulières (SVD : Singular Value Décomposition) [52] permet de donner de meilleures informations sur la structure d'une matrice. Contrairement à l'EVD, la

SVD fonctionne aussi bien sur des matrices carrées que rectangulaires et peut être appliquée directement sur X .

Soit X est une matrice réelle de $M \times N$ avec $r = \text{rang}(X) = \text{Min}(M, N)$. Il existe deux matrices orthogonales U ($UU^T = U^T U = I_M$) et V ($VV^T = V^T V = I_N$) tel que :

$$U = [u_1, \dots, u_M] \in R^{M \times M} \quad (\text{III.50})$$

$$V = [v_1, \dots, v_N] \in R^{N \times N} \quad (\text{III.51})$$

La SVD de X est donnée par :

$$X = U \Sigma V^T \quad (\text{III.52})$$

La matrice Σ , peut être partitionnée comme suit :

$$\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \quad (\text{III.53})$$

Σ_r est une matrice diagonale carrée donnée par:

$$\Sigma_r = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) \in R^{r \times r} \quad (\text{III.54})$$

Les éléments de la diagonale de Σ sont classés de telle sorte que :

- $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$ pour \hat{R}_{xx} semi-définie positive;
- $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ pour \hat{R}_{xx} définie positive.

u_i et v_i sont appelés respectivement le $i^{\text{ème}}$ vecteur singulier gauche et vecteur singulier droit de X et les éléments de la diagonale de Σ sont appelés les valeurs singulières de X .

U et V peuvent être données par :

$$U = \left[\begin{array}{c|c} \underbrace{u_1, \dots, u_r}_{\tilde{U}_r} & \underbrace{u_{r+1}, \dots, u_M}_{\tilde{U}_r} \end{array} \right] = [U_r | \tilde{U}_r] \quad (\text{III.55})$$

$$V = \left[\begin{array}{c|c} \underbrace{v_1, \dots, v_r}_{\tilde{V}_r} & \underbrace{v_{r+1}, \dots, v_N}_{\tilde{V}_r} \end{array} \right] = [V_r | \tilde{V}_r]$$

On peut représenter la SVD de X (III.52) comme suit :

$$X = [U_r | \tilde{U}_r] \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ \tilde{V}_r^T \end{bmatrix} = U_r \Sigma_r V_r^T \quad (\text{III.56})$$

SVD de X peut être également donnée par :

$$X = \sum_{i=1}^r \sigma_i u_i v_i^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T \quad (\text{III.57})$$

Les dimensions des sous-matrices sont données par :

$$U_r \in \mathbb{R}^{M \times r}, \tilde{U}_r \in \mathbb{R}^{M \times (M-r)}, V_r \in \mathbb{R}^{N \times r} \text{ et } \tilde{V}_r \in \mathbb{R}^{N \times (N-r)}.$$

Les propriétés d'orthogonalité de U et V peuvent être exprimées en fonctions des sous-matrices par :

$$\begin{aligned} U_r^T U_r &= I_r, & \tilde{U}_r^T \tilde{U}_r &= I_{M-r}, & U_r^T \tilde{U}_r &= 0, & U_r U_r^T + \tilde{U}_r \tilde{U}_r^T &= I_M \\ V_r^T V_r &= I_r, & \tilde{V}_r^T \tilde{V}_r &= I_{N-r}, & V_r^T \tilde{V}_r &= 0, & V_r V_r^T + \tilde{V}_r \tilde{V}_r^T &= I_N \end{aligned} \quad (\text{III.58})$$

On peut montrer que la SVD de X^T devient :

$$X^T = V \Sigma^T U^T = [V_r | \tilde{V}_r] \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_r^T \\ \tilde{U}_r^T \end{bmatrix} = V_r \Sigma_r U_r^T \quad (\text{III.59})$$

On peut obtenir à partir des équations (III.56), (III.58) et (III.59) ce qui suit :

$$\begin{aligned} X V_r &= U_r \Sigma_r & X v_i &= \sigma_i u_i & i &= 1, 2, \dots, r \\ X \tilde{V}_r &= 0 & X v_i &= 0 & i &= r + 1, \dots, N \\ X^T U_r &= V_r \Sigma_r & X^T u_i &= \sigma_i v_i & i &= 1, 2, \dots, r \\ X^T \tilde{U}_r &= 0 & X^T u_i &= 0 & i &= r + 1, \dots, M \end{aligned} \quad (\text{III.60})$$

(III.60) montre que u_i et v_i $i = 1, 2, \dots, r$ sont confinés dans les espaces colonnes de X et X^T désignés respectivement par $R(X)$ et $R(X^T)$.

L'équation (III.60) fournit également des bases orthonormées pour lesdits sous-espaces.

Aussi, $v_i, i = r + 1, \dots, N$ et $u_i, i = r + 1, \dots, M$ représentent des bases pour les sous-espaces nuls $N(X)$ et $N(X^T)$ respectivement.

Enfin, les matrices $U_r, \tilde{U}_r, V_r, \tilde{V}_r$ fournissent des bases orthonormées pour les quatre sous-espaces fondamentaux de X .

u_i et v_i sont désignés comme les vecteurs singuliers gauche et droit de X représentant respectivement les vecteurs propres de XX^T et $X^T X$. Cependant, il découle de l'orthogonalité de U et V que :

$$\begin{aligned} X^T X &= V \Sigma^T U^T U \Sigma V^T = V (\Sigma^T \Sigma) V^T, & \Sigma^T \Sigma &= \begin{bmatrix} \Sigma_r^2 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{N \times N} \\ XX^T &= U \Sigma V^T V \Sigma^T U^T = U (\Sigma \Sigma^T) U^T, & \Sigma \Sigma^T &= \begin{bmatrix} \Sigma_r^2 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{M \times M} \end{aligned} \quad (\text{III.61})$$

On peut remarquer à partir de (III.61) que V et U représentent les matrices des vecteurs propres de $X^T X$ et de XX^T et que les valeurs propres associées sont données par : $\lambda_i = \sigma_i^2$, $i = 1, \dots, r$. Le rang de $X^T X$ et XX^T est identique à X (r).

Nous supposons que $M \geq N$ et que le rang de la matrice X est $r = N$. La décomposition en valeurs propres (EVD) de $X^T X$ est donnée par :

$$X^T X = V \Lambda V^T, \text{ avec } \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N) \in \mathbb{R}^{N \times N} \quad (\text{III.62})$$

La matrice $X^T X$ est strictement définie positive avec des valeurs propres positives et des vecteurs propres orthonormés : $V^T V = V V^T = I_N$.

Soit $\sigma_i = \sqrt{\lambda_i}$, $i = 1, \dots, N$, et $\Sigma_1 = \Lambda^{1/2} = \text{diag}(\sigma_1, \dots, \sigma_N) \in \mathbb{R}^{N \times N}$ et $U_1 = X U_1 = X \Sigma_1^{-1}$ qui est une matrice $M \times N$ avec des colonnes orthonormées :

$$U_1^T U_1 = \Sigma_1^{-1} V^T (X^T X) V \Sigma_1^{-1} = \Sigma_1^{-1} V^T (V \Sigma_1^2 V^T) V \Sigma_1^{-1} = I_N \quad (\text{III.63})$$

Sachant que $U_1 \Sigma_1 = X V$ et $U_1 \Sigma_1 V^T = X V V^T = X$, la SVD de X devient :

$$X = U_1 \Sigma_1 V^T \quad (\text{SVD économique}) \quad (\text{III.64})$$

La matrice $U_1 (M \times N)$ peut être transformée à une matrice orthogonale $M \times M$ en ajoutant à celle-ci $M - N$ colonnes orthonormées U_2 tel que $U_2^T U_1 = 0$.

La matrice diagonale $\Sigma_1 (N \times N)$, peut être également transformée à une matrice $\Sigma (M \times N)$. On peut exprimer la SVD économique de (III.64) sous forme standard par :

$$X = U_1 \Sigma_1 V^T = [U_1 | U_2] \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} V^T = U \Sigma V^T \quad (\text{III.65})$$

L'équation (III.64) est désignée par la forme SVD économique du fait que les matrices U_1 et X possèdent les mêmes dimensions, mais avec des colonnes orthonormées et $\dim(\Sigma_1) = N \times N$.

- Approximation du rang-réduit

Les normes euclidiennes et de Frobenius de $X (M \times N)$ de rang r , peuvent être exprimées en fonction des valeurs singulières de X comme suit :

$$\begin{aligned} \|X\|_2 &= \sigma_1 = \text{valeur singulière maximale} \\ \|X\|_F &= (\sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2)^{1/2} \end{aligned} \quad (\text{III.66})$$

Nous définissons une famille de matrices de rang $l \leq r$ (rang réduit) comme suit :

$$X_l = \sum_{i=1}^l \sigma_i u_i v_i^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_l u_l v_l^T, \quad l = 1, \dots, r \quad (\text{III.67})$$

La SVD de X_l de rang réduit (III.67) peut être exprimée par :

$$X_l = U \begin{bmatrix} \Sigma_l & 0 \\ 0 & 0 \end{bmatrix} V^T, \quad \text{avec } \Sigma_l = \text{diag} \left(\sigma_1, \sigma_2, \dots, \sigma_l, \underbrace{0, \dots, 0}_{r-l \text{ zéros}} \right) \in R^{r \times r} \quad (\text{III.68})$$

Nous remarquons à partir de (III.68) que les $r - l$ valeurs singulières $(\sigma_{l+1}, \dots, \sigma_r)$ de X ont été remplacées par des valeurs nulles dans X_l .

Le théorème relatif à l'approximation de rang-réduit [30-31], indique que dans l'ensemble de $M \times N$ de rang l ($l < r$), la matrice B qui se rapproche au sens de la norme euclidienne (ou de Frobenius) le plus de X est la matrice X_l , vérifiant que la distance $\|X - B\|$ soit minimisée pour toute matrice $M \times N$ de rang l , si $B = X_l$

La distance de la matrice minimisée est donnée par :

$$\begin{aligned} \|X - X_l\|_2 &= \sigma_{l+1} \\ \|X - X_l\|_F &= (\sigma_{l+1}^2 + \sigma_{l+2}^2 + \dots + \sigma_r^2)^{1/2} \end{aligned} \quad (\text{III.69})$$

Ce théorème représente le principe de base de la représentation des données par des dimensions réduites et ce, pour les applications du rehaussement du signal, du traitement des images, etc.

Dans plusieurs applications, les valeurs singulières de la matrice X ont tendance à se regrouper en deux groupes, celles qui sont grandes et celles qui sont petites, donc en supposant que $M \geq N$, nous disposant les N valeurs singulières comme suit :

$$\underbrace{\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r}_{\text{grandes valeurs}} \gg \underbrace{\sigma_{r+1} \geq \sigma_{r+2} \geq \dots \geq \sigma_N}_{\text{petites valeurs}} \quad (\text{III.70})$$

Ce modèle est représenté sur la figure III.2.

On suppose que le rang de la matrice X est N (rang total) et que les $N - r$ valeurs singulières sont considérées comme faibles :

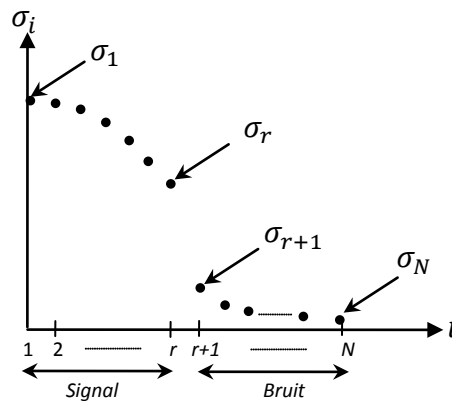


Figure III.2 : représentation des valeurs singulières dans le sous-espace du signal et sous-espace du bruit [52]

Si les $(N - r)$ dernières petites valeurs singulières dans (III.70) sont maintenues, on définit la matrice pseudo-inverse de X par :

$$X^+ = \sum_{i=1}^r \frac{1}{\sigma_i} v_i u_i^T + \sum_{i=r+1}^N \frac{1}{\sigma_i} v_i u_i^T \quad (\text{III.71})$$

et les derniers $(N - r)$ termes tendent à dominer l'expression. Pour cette raison, le rang et le pseudo-inverse de la matrice peuvent être déterminés en tenant compte d'un niveau de seuil (tolérance) δ de telle sorte que si $\sigma_i \leq \delta$ pour $i = r + 1, \dots, N$, on déduit que ces valeurs singulières peuvent être mises à zéro et que r sera efficace.

III.6. Rehaussement de la parole par Transformée de Karhunen-Loève (KLT)

La KLT [34,52] d'un vecteur aléatoire stationnaire de moyenne nulle et de dimension $(N + 1)$ $y(n) = [y_0(n), y_1(n), \dots, y_N(n)]^T$ avec matrice de covariance $R_{yy} = E[y^*(n)y^T(n)]$ est définie par la transformation linéaire suivante :

$$x(n) = Q^T y(n) \quad (\text{KLT}) \quad (\text{III.72})$$

où $Q = [q_0, q_1, \dots, q_N]$ est la matrice unitaire $(N + 1) \times (N + 1)$ des vecteurs propres de R_{yy} , à savoir :

$$R_{yy}q_i = \lambda_i q_i \quad i = 0, 1, \dots, N \quad (\text{III.73})$$

λ_i représentent les valeurs propres (supposées être dans l'ordre décroissant).

On peut également exprimer (III.73) par :

$$R_{yy}Q = Q\Lambda, \quad \Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_N\} \Rightarrow Q^+ R_{yy}Q = \Lambda \quad (\text{III.74})$$

avec Q^+ la matrice pseudo-inverse de Q .

Les éléments du vecteur $x(n) = [x_0(n), x_1(n), \dots, x_N(n)]^T$ sont appelées composantes principales. Elles peuvent être exprimées sous forme de produits entre q_i et $y(n)$ comme suit :

$$x(n) = Q^T y(n) \Rightarrow \begin{bmatrix} x_0(n) \\ x_1(n) \\ \vdots \\ x_N(n) \end{bmatrix} = \begin{bmatrix} q_0^T y_0(n) \\ q_1^T y_1(n) \\ \vdots \\ q_N^T y_N(n) \end{bmatrix}, \quad \text{où } x_i(n) = q_i^T y(n), i = 1, \dots, N \quad (\text{III.75})$$

Les composantes principales sont mutuellement orthogonales, donc non corrélées.

$Q^+ R_{yy}Q = \Lambda$ est la matrice de covariance du vecteur $x(n)$, c.à.d., :

$$E[x^*(n)x^T(n)] = Q^+ E[y^*(n)y^T(n)]Q = Q^+ R_{yy}Q = \Lambda \quad (\text{III.76})$$

où encore :

$$E[x_i^*(n)x_j^T(n)] = \lambda_i \delta_{ij}, \quad i, j = 0, 1, \dots, N \quad (\text{III.77})$$

On peut remarquer que la KLT permet la décorrélation des composantes du vecteur $y(n)$.

Les valeurs propres de R_{yy} représentent les variances des composantes principales, à savoir :

$$\sigma_i^2 = E[|x_i^2(n)|] = \lambda_i \quad (\text{III.78})$$

Etant donné que $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_N$, la composante principale $x_0(n)$ aura la plus grande variance, la deuxième composante principale $x_1(n)$ aura la deuxième plus grande variance et ainsi de suite. La variance totale de $y(n)$ est définie par la somme des variances de $(N + 1)$ composantes, où on peut remarquer qu'elle est égale à la somme des variances des Composantes Principales (CP) ou la somme des valeurs propres de R_{yy} :

$$\sigma_y^2 = \sum_{i=0}^N E[|y_i(n)|^2] \quad (\text{variance totale}) \quad (\text{III.79})$$

En utilisant la propriété de la trace ($y^+y = \text{tr}(y^*y^T)$), nous trouvons :

$$\sigma_y^2 = \text{tr}(E[y^*(n)y^T(n)]) = \text{tr}(R_{yy}) = \lambda_0 + \lambda_1 + \dots + \lambda_N \quad (\text{III.80})$$

La KLT inverse est obtenue en notant que $Q^{-T} = Q^*$ comme :

$$y(n) = Q^* x(n) \quad (\text{KLT inverse}) \quad (\text{III.81})$$

On peut exprimer (III.81) sous la forme suivante :

$$y(n) = Q^* x(n) = [q_0^*, q_1^*, \dots, q_N^*] \begin{bmatrix} x_0(n) \\ x_1(n) \\ \vdots \\ x_N(n) \end{bmatrix} = \sum_{i=0}^N q_i^* x_i(n) \quad (\text{III.82})$$

Dans beaucoup d'applications, les premières composantes principales $x_i(n)$, $0 \leq i \leq r - 1$ où $r \ll N + 1$, représente la grande partie de la variance totale, ce qui nous amène à ne garder que les r premiers termes dans la KLT inverse, à savoir :

$$\hat{y}(n) = \sum_{i=0}^{r-1} q_i^* x_i(n) \quad (\text{KLT inverse}) \quad (\text{III.83})$$

A cet effet, le signal reconstruit $\hat{y}(n)$ représente la meilleure approximation du signal original $y(n)$. L'erreur quadratique moyenne devienne :

$$E[\|y(n) - \hat{y}(n)\|^2] = E[\sum_{i=r}^N |x_i(n)|^2] = \sum_{i=r}^N \lambda_i \quad (\text{III.84})$$

III.7. Rehaussement de la parole par Analyse en Composantes Principales (PCA)

L'analyse en composantes principales (PCA) [52, 82] est une technique utilisée pour représenter des données multidimensionnelles par des données de dimensions plus petites, et ce, à travers la décomposition en valeurs propres (EVD) ou valeurs singulières (SVD) de ces données.

L'analyse PCA est aussi équivalente à KLT. La seule différence est qu'elle est appliquée, à l'inverse de KLT, directement sur la matrice de covariance simplifiée \hat{R}_{yy} .

On peut calculer \hat{R}_{yy} , à partir de M vecteurs $y(n)$, $n = 0, 1, \dots, M - 1$, comme suit :

$$\hat{R}_{yy} = \frac{1}{M} \sum_{n=0}^{M-1} y^*(n) y^T(n) \quad (\text{III.85})$$

On suppose que la moyenne des échantillons est éliminée, c.à.d., $m = \frac{1}{M} \sum_{n=0}^{M-1} y(n) = 0$.

L'équation (III.85) est simplifiée en ignorant le facteur global $\frac{1}{M}$, comme suit :

$$\hat{R}_{yy} = \sum_{n=0}^{M-1} y^*(n) y^T(n) = Y^+ Y, \quad \text{avec} \quad Y = \begin{bmatrix} y^T(0) \\ y^T(1) \\ \vdots \\ y^T(M-1) \end{bmatrix} \quad (\text{III.86})$$

où Y est la matrice de données de $M \times (N - 1)$.

On peut définir la matrice Q à partir de KLT (ou PCA) comme suit :

$$\hat{R}_{yy} Q = Q \Lambda \quad (\text{III.87})$$

Les signaux correspondants aux composantes principales sont donnés par :

$$x(n) = Q^T y(n) \Rightarrow x^T(n) = y^T(n)Q, \quad n = 0, 1, \dots, M - 1 \quad (\text{III.88})$$

(III.88) peut être exprimée, sous forme matricielle par :

$$X = YQ \quad (\text{PCA}) \quad (\text{III.89})$$

Où $X\{M \times (N + 1)\}$ est la matrice de données de $x(n)$:

$$X = \begin{bmatrix} x^T(0) \\ x^T(1) \\ \vdots \\ x^T(M - 1) \end{bmatrix} \quad (\text{III.90})$$

La transformée inverse peut être obtenue en multipliant (III.89) par Q^+ et en utilisant la propriété $XQ^+ = YQQ^+$, comme suit :

$$Y = XQ^+ \Rightarrow y(n) = Q^* x(n), \quad n = 0, 1, \dots, M - 1 \quad (\text{III.91})$$

L'équation (III.91) peut être exprimée explicitement comme suit :

$$y(n) = \sum_{i=0}^M q_i^* x_i(n), \quad n = 0, 1, \dots, M - 1 \quad (\text{III.92})$$

On peut déduire à partir de (III.87), (III.89) et (III.91) que :

$$X^+ X = Q^+ \hat{R}_{yy} Q = \Lambda \Rightarrow \sum_{n=0}^{M-1} x^*(n) x^T(n) = \Lambda \quad (\text{III.93})$$

(III.93) peut être exprimée par :

$$\sum_{n=0}^{M-1} x_i^*(n) x_j^T(n) = \lambda_i \delta_{ij}, \quad i, j = 0, 1, \dots, N \quad (\text{III.94})$$

On remarque que les composantes principales des signaux $x_i(n)$ sont identiques aux vecteurs propres singuliers gauches de la SVD de la matrice des données Y .

Supposons que les λ_i ($\lambda_i \neq 0$) désignent les valeurs singulières $\sigma_i = \sqrt{\lambda_i}$, $i = 0, 1, \dots, M - 1$.

et

$$U = X\Sigma^{-1}, \quad \Sigma = \text{diag}\{\sigma_0, \sigma_1, \dots, \sigma_N\} = \Lambda^{1/2} \quad (\text{III.95})$$

où $\dim(U) = M \times (N + 1)$, $\dim(\Sigma) = (N + 1) \times (N + 1)$.

Il s'ensuit à partir de (III.93) que U possède des colonnes orthonormées :

$$U^+U = \Sigma^{-1}X^+X\Sigma^{-1} = \Lambda^{-1/2}\Lambda\Lambda^{1/2} = I_{N+1} \quad (\text{III.96})$$

Sachant que $X = U\Sigma$ et $Y = XQ^+$ (III.91), on peut exprimer la SVD de Y comme suit :

$$Y = U\Sigma Q^+ \quad (\text{SVD économique}) \quad (\text{III.97})$$

De ce qui précède, on peut déduire que l'analyse en composantes principales basée sur \hat{R}_{yy} est équivalente à la SVD économique de la matrice de données Y .

III.8. Rehaussement de la parole par KLT-VRE

Une des principales difficultés dans le domaine du rehaussement de la parole basé sur KLT est la sélection du nombre optimal de Composantes Principales (CP), permettant de définir le modèle de reconstruction. Le problème fondamental dans le développement d'un modèle PCA est de choisir le nombre suffisant de CP pour représenter le système de façon optimale. Si le nombre de CP est choisi plus petit qu'il n'est requis, le modèle sera sous-paramétré résultant en une représentation incomplète du processus. En d'autres termes, si plus de CP qu'il n'en faut sont sélectionnés, le modèle sera sur-paramétré et comportera en conséquence du bruit.

Plusieurs méthodes sont utilisées pour calculer le nombre de CP [83-85], mais la plupart utilisent des indices croissants ou décroissants de manière monotone. Donc, la décision liée au choix du nombre de CP est très subjective. Dans cette section, nous présentons une méthode basée sur la variance de l'erreur de reconstruction (VRE) pour sélectionner le nombre de CP. Cette méthode est basée sur la recherche d'un minimum de CP. Des conditions sont données pour que ce minimum corresponde au nombre de CP nécessaires contenant le signal propre.

Dans cette partie, nous exposons quelques critères de sélection du nombre de CP tout en axant notre étude sur le critère VRE choisi dans le cadre des travaux de cette thèse.

Parmi les approches utilisées pour sélectionner le nombre optimal de CP, on peut citer les critères suivants : AIC (Akaike Information Criterion), MDL (Minimum Description Length), IEF (Imbedded Error Function) et VRE (Variance of the Reconstruction Error), développés comme suit :

III.8.1. Critères AIC et MDL

Les critères AIC et MDL ont été développés par Akaike [65] et Rissanen [75]. Leur principe de fonctionnement se résume comme suit :

Supposons que $n(k) \sim \mathcal{N}(0, \sigma_n^2)$ est une variable aléatoire indépendante et uniformément distribuée et que la matrice de covariance de $x(t)$ soit donnée par :

$$R_{xx} = E\{xx^T\} = R_{ss} + \sigma_n^2 I \quad (\text{III.98})$$

avec $\text{rang}(R_{ss}) = k \leq N$ et sachant que les valeurs propres de R_{xx} ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N \geq 0$), on peut vérifier que les $N - k$ petites valeurs propres de R_{xx} sont égales à σ_n^2 .

On aura donc la relation suivante :

$$\lambda_{k+1} = \lambda_{k+2} = \dots = \lambda_N = \sigma_n^2 \quad (\text{III.99})$$

et

$$R_{xx} = \sum_{i=1}^k (\lambda_i - \sigma_n^2) q_i q_i^T + \sigma_n^2 I \quad (\text{III.100})$$

avec $\lambda_1, \dots, \lambda_k$ et q_1, \dots, q_k représentant respectivement les k plus grandes valeurs et vecteurs propres de R_{xx} .

Etant donné le vecteur des paramètres du modèle PCA suivant [85] :

$$w = [\lambda_1, \dots, \lambda_k, \dots, \sigma_n^2, q_1^T, \dots, q_k^T]^T \in \mathcal{R}^{(N+1)k+1} \quad (\text{III.101})$$

Les critères AIC et MDL ont été initialement utilisés dans le cadre de la sélection du nombre de CP par Wax et Kailath [86].

Soit un ensemble de K observations x_1, \dots, x_K considéré comme une séquence de vecteurs Gaussien, indépendants, identiquement distribués et de moyenne nulle. On cherche à calculer l'estimateur du maximum de vraisemblance de w comme suit :

$$\hat{w} = \left[d_1, \dots, d_l, \dots, \frac{1}{N-l} \sum_{i=l+1}^N d_i, c_1^T, \dots, c_l^T \right]^T \quad (\text{III.102})$$

où d_1, \dots, d_N et c_1, \dots, c_l représentent respectivement les valeurs et vecteurs propres de la matrice de covariance \hat{R}_{xx} définie dans (III.5).

Il a été constaté [86] que les fonctions AIC et MDL ont la forme suivante :

$$\begin{aligned} AIC(l) &= -2\log\left(\frac{\prod_{s=l+1}^N d_s^{1/(N-1)}}{\frac{1}{N-1}\sum_{s=l+1}^N d_s}\right)^{(N-1)K} + 2T \\ MDL(l) &= -2\log\left(\frac{\prod_{s=l+1}^N d_s^{1/(N-1)}}{\frac{1}{N-1}\sum_{s=l+1}^N d_s}\right)^{(N-1)K} + T\log K \end{aligned} \quad (\text{III.103})$$

où T est le nombre de paramètres indépendants dans \hat{w} .

$$T = \begin{cases} l(2N - l) & \text{si } c_1, \dots, c_l \text{ complexes} \\ \frac{1}{2}(N - l + 1) & \text{si } c_1, \dots, c_l \text{ réels (cas du signal de parole)} \end{cases} \quad (\text{III.104})$$

Le premier terme de $AIC(l)$ ou $MDL(l)$ tend à diminuer et le second terme augmente avec l entraînant une valeur minimale pour $l \in [1, N]$. Wax et Kailath [86] ont montré que, pour une valeur de K assez grande, le critère $MDL(l)$ donne une estimation cohérente, tandis que $AIC(l)$ tend à surestimer le nombre de PC.

III.8.2. Critère IEF

L'idée de base de la méthode IEF [87] est que l'espace de mesure peut être divisé en deux sous-espaces: un sous-espace de composante principale qui représente le signal propre et un sous-espace résiduel contenant le bruit. La fonction IEF peut être calculée par rapport aux valeurs propres comme suit :

$$IEF(l) = \left[\frac{l \sum_{j=l+1}^N \lambda_j}{KN(N-1)} \right]^{1/2} \quad (\text{III. 105})$$

où l représente le nombre de CP utilisé pour représenter les données. Ainsi, chaque composante principale contient une portion du signal et une portion du bruit. Avant l'extraction totale de tous les signaux, $IEF(l)$ contient un mélange de bruit et du signal, qui

tendent à diminuer avec l . Lorsque tous les signaux sont extraits, il reste uniquement du bruit qui tend à augmenter avec l . La valeur minimale de $IEF(l)$ correspond donc au nombre de CP nécessaires pour reconstituer les données.

III.8.3. Critère VRE

Les critères de sélection du nombre de CP utilisés précédemment sont subjectifs. D'autre part, la plupart d'entre eux utilisent un index qui décroît de façon monotone. Le nombre de CP est choisi de sorte qu'il n'y a pas de diminution significative de l'indice après l'ajout d'une CP. Ces approches basées sur des index monotones sont subjectives au sens où : (i) il peut y avoir une diminution assez constante de l'indice, et : (ii) il ne peut y avoir plus d'un emplacement qui satisfait au critère. Cependant, le nombre de CP a un impact significatif sur le rehaussement du signal de la parole, au sens où si le nombre de CP est choisi plus petit qu'il n'est requis, le modèle sera sous-paramétré résultant en une représentation incomplète du processus. Par ailleurs, si plus de CP qu'il n'en faut sont sélectionnés, le modèle sera sur-paramétré et comportera en conséquence du bruit.

Qin et Dunia [84] ont proposé un nouveau critère de sélection du nombre de CP basé sur la meilleure reconstruction des variables. Une caractéristique importante de cette approche est que l'indice proposé a un minimum (non-monotone) correspondant à la meilleure reconstruction. Quand le modèle de PCA est utilisé pour reconstruire le signal bruité, l'erreur de reconstruction est une fonction du nombre de CP. Ces auteurs ont donc proposé l'utilisation de la VRE pour déterminer le nombre de CP. La VRE proposée peut être décomposée en une partie du sous-espace des CP (PCS) et une partie dans le sous-espace résiduel (RS). La partie dans RS décroît de façon monotone avec le nombre de CP et celle dans (PCS) augmente avec le nombre de PC impliquant que la VRE dispose toujours d'un minimum qui coïncide avec le nombre optimal de CP.

Etant donné le modèle PCA décrit précédemment. On procède à la séparation des vecteurs propres et des matrices des CP comme suit :

$$Q = \{\hat{Q}_l | \tilde{Q}_{N-l}\}, \quad Y = \{\hat{Y}_l | \tilde{Y}_{N-l}\} \quad (\text{III.106})$$

$$X = \hat{Y}_l \hat{Q}_l^T + \tilde{Y}_{N-l} \tilde{Q}_{N-l}^T = \hat{X} + N \quad (\text{III.107})$$

avec

$$\hat{X} = X\hat{C}_l \quad \text{et} \quad N = X\tilde{C}_{N-l} \quad (\text{III.108})$$

Le modèle PCA peut être constitué par les matrices : $\hat{C}_l = \hat{Q}_l\hat{Q}_l^T$ et $\tilde{C}_{N-l} = I_N - \hat{C}_l$.

$l \{ l \leq r \}$ représente le nombre optimal de CP utilisées pour la reconstruction du signal.

Qin et Dunia [84] ont proposé de déterminer l par minimisation de la variance de l'erreur de reconstruction. La reconstruction des variables consiste à les estimer en utilisant le modèle PCA.

Nous supposons que le signal est corrompu avec un bruit n_j le long d'une direction $\xi_j \in \mathbb{R}^N$:

$$x = s + n_j\xi_j, \quad (\text{III.109})$$

où x désigne le signal bruité, s le signal propre inconnu et $\|\xi_j\| = 1$.

ξ_j , correspond à la $j^{\text{ième}}$ colonne de la matrice identité.

L'opération de reconstruction du signal consiste à trouver une estimée de s le long de la direction ξ_j du bruit pour mieux corriger l'effet de ce dernier. En d'autres termes, nous essayons de trouver \hat{n}_j tels que:

$$\hat{s} = x - \hat{n}_j\xi_j \quad (\text{III.110})$$

Le modèle de l'erreur minimale est:

$$n_j = \underset{n_j}{\operatorname{argmin}} \|s - \hat{s}\|^2 = \underset{n_j}{\operatorname{argmin}} \|\tilde{s}\|^2 = \underset{n_j}{\operatorname{argmin}} \|\tilde{x} - \hat{n}_j\tilde{\xi}_j\|^2 \quad (\text{III.111})$$

\hat{s} et \tilde{s} désignent respectivement la portion du signal propre dans le sous-espace des CP (PCS) et le sous-espace résiduel(RS).

La solution de l'équation (III.111) peut être trouvée en utilisant l'estimateur des moindres carrés LS [32] :

$$\hat{n}_{j,LS} = (\xi_j^T \xi_j)^{-1} \xi_j^T \tilde{x} = \frac{\xi_j^T \tilde{x}}{\xi_j^T \xi_j} \quad (\text{III.112})$$

Nous pouvons obtenir la meilleure reconstruction en substituant $\hat{n}_{j,LS}$ dans (III.110) comme suit :

$$\hat{s} = \left(I - \frac{\tilde{\xi}_j \tilde{\xi}_j^T}{\tilde{\xi}_j^T \tilde{\xi}_j} \right) x = (-\tilde{\xi}_j^0 \tilde{\xi}_j^{0T}) x \quad (\text{III.113})$$

où $\tilde{\xi}_j^0 \equiv \tilde{\xi}_j / \|\tilde{\xi}_j\|$.

Dans le cas du signal bruité, $\tilde{\xi}_j$ est la $j^{\text{ième}}$ colonne de la matrice identité, c.-à-d., $\tilde{\xi}_j = [0 \dots 1 \dots 0]^T$.

Par exemple : $\tilde{\xi}_j = [1 \ 0 \ 0 \ 0 \ 0]^T$ signifie que parmi les cinq dimensions, seulement la première dimension est bruitée. A cet effet, l'équation (III.113) fournit la meilleure reconstruction basée sur l'analyse PCA du signal bruité.

En substituant l'équation (III.109) dans (III.110), on obtient :

$$s - \hat{s} = (\hat{n}_j - n_j) \tilde{\xi}_j \quad (\text{III.114})$$

En substituant l'équation (III.109) dans (III.111) on obtient :

$$\hat{n}_j - n_j = \frac{\tilde{\xi}_j^T s}{\tilde{\xi}_j^T \tilde{\xi}_j} \quad (\text{III.115})$$

$$s - \hat{s} = (\hat{n}_j - n_j) \tilde{\xi}_j = \frac{\tilde{\xi}_j^T s}{\tilde{\xi}_j^T \tilde{\xi}_j} \tilde{\xi}_j \quad (\text{III.116})$$

$$\|s - \hat{s}\| = \frac{|\tilde{\xi}_j^T s|}{\tilde{\xi}_j^T \tilde{\xi}_j} \quad (\text{III.117})$$

Nous pouvons observer à partir des équations (III.115) et (III.116) que :

- $E\{s - \hat{s}\} = 0$, parce que $E\{s\} = 0$. En d'autres termes, \hat{s} est un estimateur non biaisé de s ;
- La variance de $(s - \hat{s})$ se produit seulement dans la direction de reconstruction $\tilde{\xi}_j$, ce qui représente le même cas pour $(n_j - \hat{n}_j)$;
- L'erreur de reconstruction $(s - \hat{s})$ est indépendante de n_j ;
- L'erreur de reconstruction $(s - \hat{s})$ dépend du nombre de CP retenu dans le modèle PCA.

Les deux derniers points indiquent qu'on peut déterminer le nombre optimal de CP en réalisant l'erreur de reconstruction minimale. En outre, si l'erreur de reconstruction est minimisée pour une amplitude particulière n_j , elle l'est pour toutes les autres amplitudes.

Par conséquent, une meilleure reconstruction du signal peut être obtenue en traitant seulement le cas $n_j = 0$.

Supposons $n_j = 0$, la VRE dans la direction ξ_j peut être calculée comme suit [84] :

$$\mu_j(l) = \text{var}\{\xi_j^T(x - \hat{s})\} = \text{var}\{\hat{n}_j\} = \frac{\tilde{\xi}_j^T(l)\hat{R}_{xx}\tilde{\xi}_j(l)}{\left(\tilde{\xi}_j^T(l)\tilde{\xi}_j(l)\right)^2} \quad (\text{III.118})$$

où

$$\tilde{\xi}_j(l) = \left(I - \hat{Q}(l)\hat{Q}^T(l)\right)\xi_j \quad (\text{III.119})$$

\hat{R}_{xx} désigne la matrice de covariance calculée à partir des équations (III.4) et (III.5). μ_j représente la variance de l'erreur de reconstruction dans l'estimation de s en utilisant \hat{s} . $\hat{Q}(l)$ est obtenue à partir de (III.11) en gardant uniquement les l premières colonnes.

Dans le but de trouver le nombre de CP, μ_j doit être minimisée par rapport au nombre de CP. Considérons différentes directions du bruit et supposons μ_j pour toutes les directions, la VRE peut être donnée par :

$$VRE(l) = \sum_{j=1}^N \frac{\mu_j(l)}{\text{var}\{\xi_j^T x\}} = \sum_{j=1}^N \frac{\mu_j(l)}{\xi_j^T \hat{R}_{xx} \xi_j} \quad (\text{III.120})$$

l est le rang de la matrice du signal propre qui minimise (III.118).

Enfin, la sélection du nombre de CP peut être effectuée en suivant les étapes suivantes :

- Construction du modèle PCA à partir des données bruitées;
- Calcul de μ_j et de la VRE en utilisant (III.118) et (III.120) ;
- Le minimum de la VRE apparaît pour un nombre spécifique de CP qui correspond à la meilleure reconstruction.

Afin de reconstruire un signal à travers une observation, il est nécessaire d'identifier en premier lieu le modèle du signal de parole et d'enlever, en deuxième lieu, le sous-espace du bruit seulement, suivi de la modification du sous-espace du signal en vue d'éliminer l'effet de bruit.

Nous utilisons l'estimateur TDC pour dériver le signal rehaussé comme suit :

$$\hat{S} = \hat{Q}_l G_1 \hat{Q}_l^T X = H_{TDC} X \quad (\text{III.121})$$

où \hat{Q}_l est la matrice des vecteurs propres tronquée, obtenue en éliminant les dernières $N - l$ colonnes de Q . L'estimateur linéaire peut être exprimé par :

$$H_{TDC} = \sum_{k=1}^l g_1(k) q_k q_k^T \quad (\text{III.122})$$

où $g_1(k)$ désigne le $k^{\text{ième}}$ élément diagonal de G_1 donné par :

$$G_1 = \sum_{k=1}^l g_1(k) = \sum_{k=1}^l \frac{\lambda_s(k)}{\lambda_s(k) + \mu \sigma_n^2} \quad (\text{III.123})$$

III.9. Conclusion

Ce chapitre a été consacré aux méthodes de rehaussement de la parole basées sur la décomposition en sous espace du signal. Quatre méthodes ont été abordées à savoir : la décomposition en valeurs singulières (SVD), la décomposition en valeurs et vecteurs propres (EVD), la transformée de Karhunen-Loève (KLT) et l'analyse en composantes principales (PCA). Nous avons vu que ces transformations présentent des similitudes et que la différence réside particulièrement sur le choix de la matrice de covariance.

Dans le cas de la PCA et la KLT, la différence concerne l'utilisation dans PCA de la matrice de covariance estimée \hat{R}_{yy} et dans le décalage par rapport à l'origine d'une distance correspondant à la moyenne du vecteur bruité $y(n)$.

Aussi, nous avons montré que les CP des signaux $x_i(n)$ sont identiques aux vecteurs propres singuliers gauches de la SVD de la matrice des données Y .

Contrairement à l'EVD, la SVD fonctionne aussi bien sur des matrices carrées que rectangulaires et peut être appliquée directement sur la matrice des données.

Il a été également déduit que l'analyse PCA basée sur \hat{R}_{yy} est équivalente à la SVD économique de la matrice de données Y .

A la fin de ce chapitre, il a été abordé les critères de sélection du nombre optimal de CP ainsi que le choix de la VRE qui permet une meilleure reconstruction des variables.

Le prochain chapitre sera consacré à la présentation de deux nouvelles méthodes de rehaussement de la parole proposées, basées sur la décomposition en sous-espaces du signal et sur le modèle perceptuel en les associant avec le critère VRE et l'estimation continue de la DSP du bruit (MCRA : Minima Controlled Recursive Averaging).

Chapitre IV

Méthodes PKLT-VRE et KLT-VRE-MCRA de rehaussement de la parole proposées

IV.1. Introduction

Ce chapitre est consacré à la présentation de deux nouvelles méthodes de décomposition en sous-espaces basées sur l'optimisation des paramètres par le biais d'une fonction de suppression permettant une meilleure reconstruction des variables. Dans les méthodes de décomposition en sous-espaces de l'état de l'art, l'optimisation des paramètres KLT a été effectuée à travers l'estimateur MMSE qui donne une meilleure reconstruction du signal propre pour des bruits stationnaires. Cependant, la sélection de l'ordre du modèle KLT ainsi que la reconstruction des variables demeurent des tâches difficiles dans le cas d'un bruit non stationnaire.

L'optimisation des paramètres KLT est également liée au contexte environnemental qui change de manière imprévisible. Une réduction importante de la dimension du modèle KLT peut engendrer une perte considérable de l'information, tandis qu'un modèle de reconstruction avec un ordre supérieur de paramètres KLT sera considéré comme insuffisant pour éliminer le bruit.

La première méthode proposée dénommée KLT-VRE-MCRA consiste à combiner avec le modèle KLT, le critère VRE et la méthode MCRA destinés respectivement à la sélection du nombre optimal de Composantes Principales à retenir pour la reconstruction du signal propre et à l'estimation continue de la DSP du bruit à travers la poursuite des variations brusques des bruits non stationnaires. Trois méthodes d'estimation du bruit ont été combinées avec le modèle KLT-VRE, à savoir : MCRA, MCRA2 et IMCRA (Improved MCRA).

La deuxième méthode proposée dénommée PKLT-VRE concerne d'une part, l'optimisation des paramètres KLT par le critère VRE et d'autre part la proposition d'une nouvelle matrice de gain de suppression qui a été ensuite optimisée par une fonction de régression linéaire.

IV.2. Définition des paramètres communs

IV.2.1. Corpus de tests

Le corpus de phrases que nous avons utilisé est issu de la base de données NOIZEUS [60] qui contient 30 phrases phonétiquement équilibrées, prélevées de celui d'IEEE [88], prononcées par 3 locuteurs masculins et 3 locutrices féminines. Ces phrases ont été corrompues par huit différents types de bruits issus de la base de données AURORA (Train, Chahut, Bus, Salle d'expositions, Restaurant, Rue, aéroport et Gare) à des RSB de 0,5, 10, 15dB. Ces fichiers ont été ensuite enregistrés au format WAV. Le choix de la base de données NOIZEUS est motivé par le fait qu'elle a été initialement développée pour les besoins du rehaussement de la parole, à l'effet de permettre aux différents chercheurs de disposer, dans des conditions similaires, d'une même plateforme de comparaison de leurs éventuels algorithmes de rehaussement.

IV.2.2. Evaluation des performances

- Mesures subjectives

La mesure subjective de la parole la plus utilisée est désignée par la méthode du jugement qui permet à des auditeurs d'évaluer la qualité du signal de test à l'aide d'une échelle de 1 à 5. Le résultat final de cette mesure est obtenu en moyennant les scores affichés par tous les auditeurs. Ce moyennage des scores est communément désigné par le Score Moyen d'Opinions (MOS : Mean Opinion Score). Les tests d'écoute subjectifs peuvent être considérés comme le moyen le plus fiable d'évaluation de la qualité de la parole. Cependant, ces tests peuvent prendre beaucoup de temps et nécessitant dans la plupart des cas des auditeurs qualifiés. Pour cela et en vue de réduire l'incertitude de l'auditeur, souvent appelé à assister à des tests successifs, il a été proposé dans [89] une méthode numérique qui se substitue à la méthode de jugement classique et qui permet d'évaluer le signal de la parole propre (SIG), le bruit (BACK) ainsi que la qualité globale de la parole (OVL). Ces mesures sont représentées comme suit :

- **SIG** **5**=très naturel non dégradé, **4**=assez naturel, faible dégradation, **3**= légèrement naturel, légèrement dégradé, **2**= assez peu naturelle, assez dégradé, **1**= très peu naturel, très peu dégradé ;
- **BAK** **5**= imperceptible, **4**= légèrement perceptible, **3**= faiblement perceptible, **2**= assez visible, **1**= très visible ;
- **OVL** **5**=Excellent (imperceptible), **4**=Bon (à peine perceptible, mais pas gênant),
(MOS)

3= Moyen (perceptible, et un peu gênant), **2=** Faible (gênant, mais pas désagréable), **1=** Mauvais (très gênant et désagréable).

- Mesures objectives

Les mesures objectives devront incorporer des connaissances psychoacoustiques, prosodiques, sémantiques, linguistiques et pragmatiques [1]. Elles permettent d'évaluer la distorsion entre le signal propre et le signal rehaussé.

Dans le cadre de cette thèse, nous avons opté pour les mesures objectives suivantes :

- Rapport Signal sur Bruit segmental (SegSNR : Segmental Signal to Noise Ratio) [90] ;
- Evaluation Perceptive de la Qualité de la Parole (PESQ : Perceptual Evaluation of Speech Quality) [91] ;
- Pente Spectrale Pondérée (WSS : Weighted Slope Spectral) [92] ;
- Logarithme du Rapport de Vraisemblance (LLR : Log Likelihood Ratio) [93].

IV.3. Méthode de rehaussement de la parole "KLT-VRE-MCRA" proposée

Dans tous les algorithmes de rehaussement de la parole, l'estimation du bruit peut avoir un impact majeur sur la qualité du signal rehaussé. Si l'estimation du bruit est trop basse, le bruit résiduel (bruit musical) deviendra audible et si le niveau d'estimation du bruit est trop élevé, la valeur estimée du signal propre deviendra distordue. En effet, le bruit peut changer rapidement d'une composante spectrale à une autre, ce qui induit à un RSB différent. Typiquement, la DSP du bruit peut être estimée par le biais d'un détecteur d'activité vocale (VAD) [59] durant la période de silence. Cependant, une estimation beaucoup plus fine de la DSP du bruit peut être obtenue si nous procédons de façon continue. Pour cela, nous allons utiliser la méthode MCRA qui consiste à traquer le minimum de la puissance du signal bruité dans chaque bande de fréquence.

IV.3.1. Description de la méthode MCRA

Nous utilisons trois types de MCRA, à savoir : MCRA [13], MCRA2 [14] et IMCRA [15].

- La méthode MCRA consiste à moyenniser les valeurs précédentes du spectre de puissance, en utilisant un paramètre de lissage qui est ajusté par la probabilité de présence du signal dans les différentes sous-bandes.
- La méthode MCRA2 consiste à mettre à jour l'estimée du bruit par un moyennage du spectre de puissance du signal bruité en utilisant des facteurs de lissage dépendant du temps

et de la fréquence et qui sont ajustés par la probabilité de présence du signal dans chaque bande.

- La méthode IMCRA (Improved MCRA) consiste à estimer les valeurs moyennes du spectre de puissance du bruit en utilisant un paramètre de lissage dépendant du temps et de la fréquence et qui est ajusté par la probabilité de présence du signal.

La forme générale de l'algorithme d'estimation récursive du bruit peut être représentée par [13] :

$$\sigma_n^2(\mathcal{L}, k) = \alpha_n(\mathcal{L}, k)\sigma_n^2(\mathcal{L} - 1, k) + [1 - \alpha_n(\mathcal{L}, k)]|X(\mathcal{L}, k)|^2 \quad (\text{IV.1})$$

Où $|X(\mathcal{L}, k)|^2$ et $\sigma_n^2(\mathcal{L}, k)$ désignent respectivement le carré du spectre d'amplitude du signal bruité et la valeur estimée de la DSP du bruit relatifs à la trame \mathcal{L} et la fréquence k .

$\alpha_n(\mathcal{L}, k)$ est un facteur de lissage à temps variable, donné par :

$$\alpha_n(\mathcal{L}, k) = \alpha + (1 - \alpha)p(\mathcal{L}, k), \quad 0 \leq \alpha \leq 1 \quad (\text{IV.2})$$

$p(\mathcal{L}, k) \triangleq P(H_1^k(\mathcal{L}, k))$ représente la probabilité conditionnelle de présence de la parole pouvant être calculée par le rapport entre la puissance du signal bruité et son minimum tel qu'il est décrit ci-dessous.

Le processus d'implémentation de la méthode MCRA peut être décrit comme suit :

1. Calcul de la DSP lissée du signal bruité $X(\mathcal{L}, k)$;
2. Calcul du rapport entre l'énergie locale du signal bruité et son minimum :

$$X_r(\mathcal{L}, k) = \frac{X(\mathcal{L}, k)}{X_{min}(\mathcal{L}, k)} \quad (\text{IV.3})$$

3. Calcul de la probabilité de présence de la parole $p(\mathcal{L}, k)$ en utilisant la récursion du premier ordre suivante :

$$p(\mathcal{L}, k) = \alpha_p p(\mathcal{L} - 1, k) + (1 - \alpha_p)I(\mathcal{L}, k) \quad (\text{IV.4})$$

avec $I(\mathcal{L}, k) = \begin{cases} 1, & X_r(\mathcal{L}, k) > \delta \\ 0, & \text{autrement} \end{cases}$

δ est un seuil qui peut avoir une valeur constante ou variable suivant le type de méthode MCRA à utiliser.

4. Calcul du facteur de lissage $\alpha_n(\mathcal{L}, k)$ (IV.2) ;
5. Mise à jour de la valeur estimée du bruit $\sigma_n^2(\mathcal{L}, k)$ (IV.1) ;

IV.3.2. Description de la méthode "KLT-VRE-MCRA"

La figure IV.1, illustre les étapes d'implémentation de la méthode KLT-VRE-MCRA [50] proposée comme suit :

1. **Matrice de Hankel** : Arrangement du vecteur d'observation de K dimensions en matrice d'observations $X(M \times N)$ sous forme de Hankel ;
2. **Matrice de Covariance** : Calcul de la matrice de covariance du signal bruité \hat{R}_{xx} ;
3. **Décomposition en valeurs propres (EVD)** : Décomposition de \hat{R}_{xx} en valeurs et vecteurs propres ;
4. **Variance de l'erreur de reconstruction (VRE)** : Calcul du nombre l des Composantes Principales ;
5. **Matrices tronquées des valeurs et vecteurs propres** : Constitution en fonction de l des nouvelles matrices tronquées des valeurs et vecteurs propres $\hat{Q}(l), \hat{Q}^T(l)$ et $\Lambda(l)$, $\hat{Q}^T(l)$ et $\Lambda(l)$;
6. **Estimation du bruit** : Estimation de la DSP du bruit σ_n^2 par MCRA ;
7. **Nouvelle matrice du gain** : Calcul de la nouvelle matrice du gain de suppression, utilisée pour la reconstruction du signal propre ;
8. **Reconstruction du signal** : Constitution du filtre optimal H_{TDC} pour la reconstruction du signal $\hat{S} = H_{TDC} X$ (III.121);
9. **IKLT** : Extraction du signal rehaussé $\hat{s}(t)$.

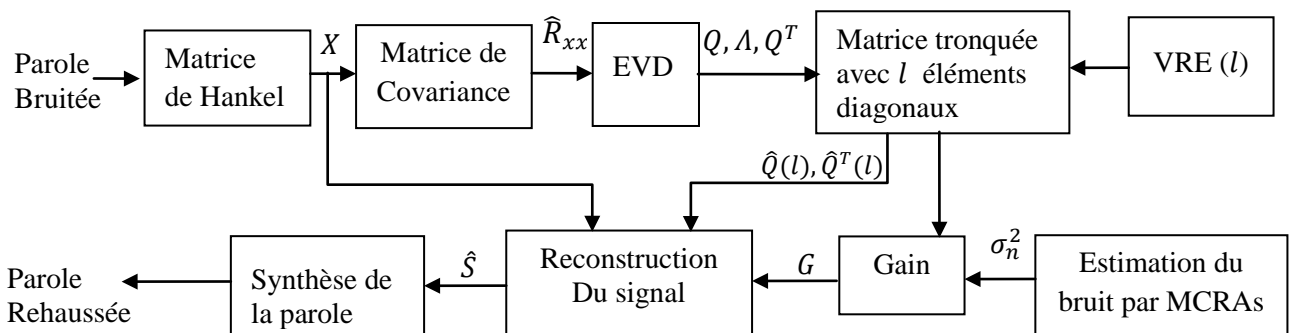


Figure IV.1 : Schéma bloc de la méthode KLT-VRE-MCRA proposée [50]

IV.3.3. Résultats et discussions

Nous avons extrait à partir de la base de données NOIZEUS, 20 phrases mélangées avec des bruits de chahut (babble noise), de rue (street noise) et d'aéroport (airport noise) à des niveaux de RSB de 0, 5 et 10 dB. Le signal bruité est ensuite échantillonné à 8KHz puis segmenté, à travers la fenêtre de Hamming, en plusieurs trames de 30 ms avec un recouvrement de 40%.

L'évaluation de la méthode KLT-VRE-MCRA proposée a été effectuée par les mesures objectives de types PESQ et RSB (SNR). La robustesse de la méthode proposées a été confirmée par les résultats obtenus, illustrés dans le tableau IV.1, qui ont été comparés avec ceux issus des méthodes de rehaussement suivantes : KLT-VRE [48,54, 94-96], Wiener [59], Logarithme de l'Amplitude Spectrale (LSA) [23] et Soustraction Spectrale(SS) [18].

Les paramètres d'optimisation de ces algorithmes ont été choisis comme suit :

Dans KLT-VRE, $N = 21$. Dans LSA, le facteur de pondération utilisé dans l'approche Décision-Directe est $a = 0.98$. Dans Wiener, le facteur de lissage pour l'approche de Décision-Directe et le facteur de lissage utilisé pour la mise à jour du bruit sont 0.99. Dans SS, le facteur d'échelle dans la période de silence est de 0.003 et le facteur de lissage utilisé pour la mise à jour du bruit est de 0.99. Dans les approches MCRA, MCRA2 et IMCRA, la valeur du seuil est choisie constante ($\delta = 5$) dans [13], et dépendante de la fréquence ($\delta(k)$) dans [14-15].

On peut remarquer à partir du tableau IV.1 que la méthode KLT-VRE-MCRA présente des résultats meilleurs par rapport aux autres méthodes utilisées pour les tests, à l'exception des cas suivants :

- (i) pour le bruit de rue, les valeurs du PESQ et RSB des méthodes KLT-VRE et Wiener sont plus grandes,
- (ii) pour le bruit d'aéroport à 0dB, la valeur du PESQ de la méthode KLT-VRE dépasse les autres.

La variante KLT-VRE-IMCRA est également légèrement plus robuste par rapport à KLT-VRE-MCRA et KLT-VRE-MCRA2, à l'exception des cas de bruits de chahut et de rue à 0 et 5 dB (RSB) et d'aéroport à 0dB (PESQ).

Aussi, nous avons remarqué que sur 50% des phrases testées, la variante KLT-VRE-IMCRA est plus performante que les deux variantes KLT-VRE-MCRA et KLT-VRE-MCRA2. Ces

dernières, avec les mêmes performances, ont présenté notamment des résultats meilleurs sur les 28% des phrases restantes.

Les résultats de cette évaluation peuvent être confirmés par les spectrogrammes de la figure IV.2.

Tableau IV.1 : Evaluations objectives (PESQ et RSB) de la méthode KLT-VRE-MCRA comparées avec les méthodes de Wiener, SS, LSA et KLT-VRE.

Type de bruit	Mesures Objective	RSB d'entrée (dB)	Wiener	SS	LSA	KLT-VRE	KLT-VRE-MCRAs		
							KLT-VRE-MCRA	KLT-VRE-MCRA2	KLT-VRE-IMCRA
Babble	RSB de sortie (dB)	0	3.55	3.29	3.03	4.47	5.19	5.19	5.77
		5	6.38	5.77	6.69	7.14	7.75	7.75	6.93
		10	11.91	10.28	11.65	11.77	11.89	11.89	12.30
	PESQ	0	0.85	0.92	1.64	1.53	1.89	1.78	1.94
		5	1.86	1.53	1.85	1.89	2.35	2.35	2.68
		10	3.93	1.81	2.65	3.85	3.98	3.98	4.21
Street	RSB de sortie (dB)	0	3.68	3.15	3.42	4.11	5.33	5.33	3.22
		5	5.27	6.87	6.82	7.09	7.47	7.47	7.33
		10	11.83	9.65	10.44	10.53	10.55	10.55	9.5
	PESQ	0	1.58	1.47	1.88	2.21	2.25	2.25	2.47
		5	2.33	1.52	2.02	2.68	2.42	2.42	2.61
		10	3.12	2.62	2.84	3.54	3.73	3.73	4.13
Airport	RSB de sortie (dB)	0	1.92	2.32	2.64	2.68	3.22	3.22	3.59
		5	4.28	4.93	5.33	5.58	7.33	7.33	8.04
		10	8.54	9.94	10.18	11.83	12.75	12.75	11.30
	PESQ	0	1.42	1.51	1.71	2.29	2.17	2.17	2.19
		5	1.77	1.87	1.92	2.41	2.38	2.38	2.67
		10	2.52	2.75	2.77	3.09	3.12	3.12	3.81

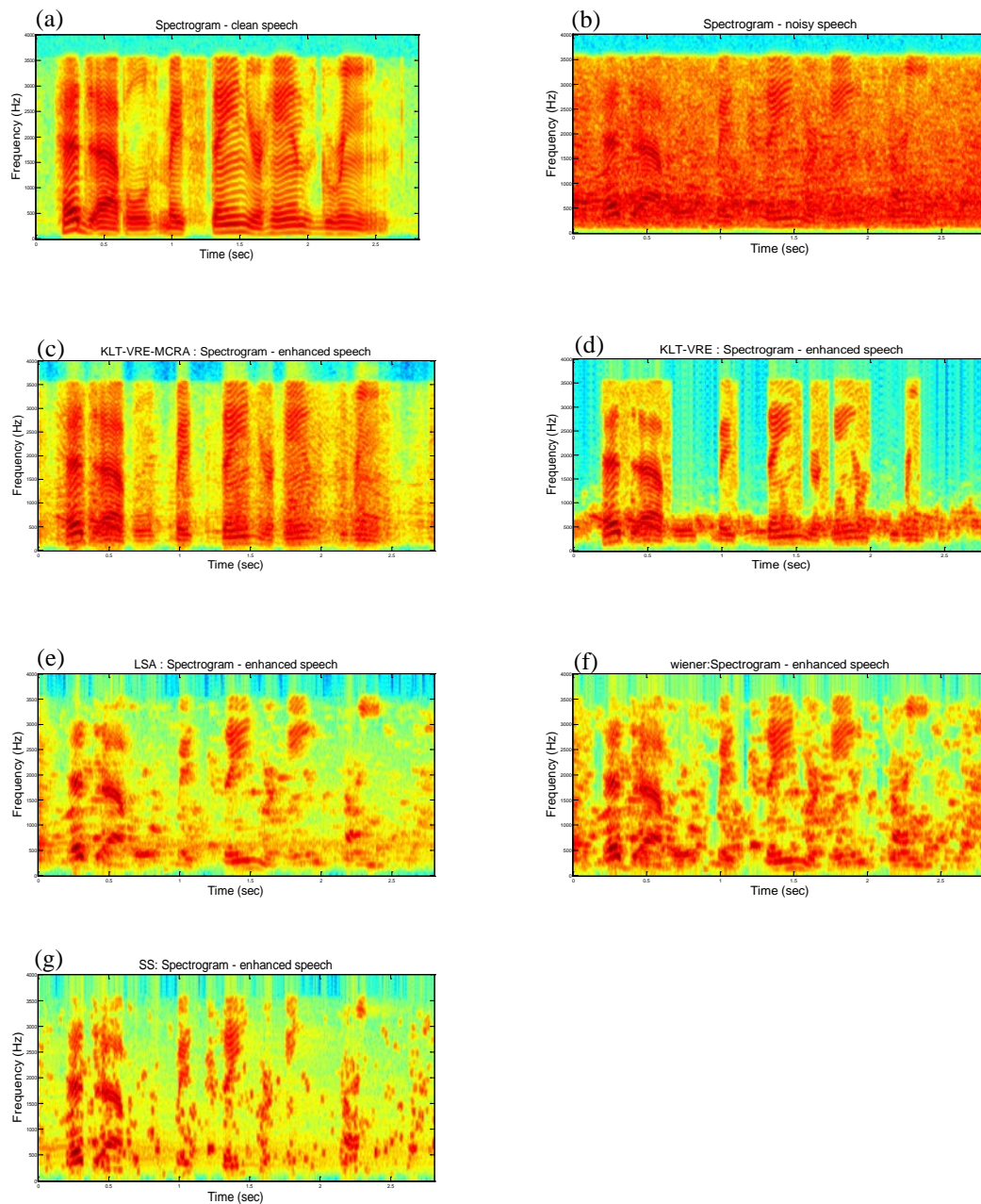


Figure IV.2 : Spectrogramme de la phrase " *Hedge apples may stain your hands green* " issue de la base de données NOIZEUS, corrompue par un bruit de type chahut (babble noise) : (a) parole propre (b) parole bruitée, (c) rehaussement par la méthode KLT-VRE-MCRA proposée, (d) rehaussement par la méthode KLT-VRE, (e) rehaussement par la méthode LSA, (f) rehaussement par la méthode de Wiener et (g) rehaussement par la méthode SS.

IV.4. Méthode de rehaussement de la parole "PKLT-VRE" proposée

Les propriétés psychoacoustiques du système auditif, incorporant le seuil de masquage de l'oreille interne, ont été utilisées dans les méthodes de rehaussement de la parole en sous-espace pour dériver un seuil de masquage spectral à court terme permettant l'estimation du niveau maximum du bruit additif pouvant être inaudible. Le phénomène du masquage peut être expliqué par les bandes critiques. A l'intérieur d'une bande critique, un son (masqué) devient inaudible en présence d'un autre bruit (masquant) avec une intensité plus élevée. Les propriétés du masquage, telles qu'elles ont été initialement présentées dans [41], sont liées aux bandes critiques. Celles-ci ont été utilisées dans le modèle KLT par le biais d'une transformation fréquence – domaine propre ou FET (Frequency Eigen-domain Transformation).

Il a été démontré dans [42] que les valeurs propres λ_i , peuvent être exprimées par :

$$\lambda_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi(\omega) |V_i(\omega)|^2 d\omega, \quad i = 1, \dots, N, \quad (IV.5)$$

avec
$$V_i(\omega) = \sum_{k=0}^{K-1} q_i(k) e^{-j\omega k}, \quad i = 1, \dots, N \quad (IV.6)$$

où $q_i = [q_i(0), \dots, q_i(K-1)]^T$ correspond aux vecteurs propres de \hat{R}_{xx} . $V_i(\omega)$, $\Phi(\omega)$ et $\Phi_B(\omega)$ représentent respectivement la Transformée de Fourier Discrète des vecteurs propres $q_i(k)$, la DSP du signal bruité calculée à partir du périodogramme ainsi que la DSP du signal bruité calculée à partir de l'estimateur de Blackman-Tukey [42].

λ_i définit une relation entre le domaine des Transformées de Fourier et de Karhunen-Loève.

La méthode PKLT [42] peut être décrite par les étapes suivantes :

1. **Estimation du bruit** : Calcul de la DSP du bruit $\Phi_n(\omega)$ et de la matrice de covariance du bruit \hat{R}_{nn} ;
2. **Calcul du sous-espace du signal** : Estimation de la matrice de covariance du signal propre $\hat{R}_{ss} = \hat{R}_{xx} - \hat{R}_{nn}$ et décomposition en valeurs propres (EVD) de \hat{R}_{ss} pour l'extraction des valeurs propres $\lambda = [\lambda_1, \dots, \lambda_N]^T$, de la matrice des vecteurs propres Q_1 et de la matrice V ;
3. **Calcul du seuil de masquage** : Utilisation de la mesure perceptuelle (Bark) pour relier la fréquence acoustique à une résolution non linéaire de la fréquence perceptuelle telle que,

1Bark couvre une bande critique. L'expression analytique permettant de réaliser le lien entre la fréquence f (Hz) à une bande critique z (Bark) est donnée par :

$$z(f) = 13 \arctan(0.00076f) + 35 \arctan \left[\left(\frac{f}{7500} \right)^2 \right] \quad (\text{IV.7})$$

À partir de (IV.7), on peut calculer le seuil de masquage pour chaque composante. La composante masquée à une fréquence particulière est rejetée si elle est en dessous d'un seuil absolu d'écoute relatif à cette fréquence. Nous déterminons le seuil de masquage d'un son avec tonalité ou sans tonalité comme suit:

$$\Phi_{\text{th}}(j, i) = X(j) + O(j) + SF(j, i) \quad (\text{IV.8})$$

où $\Phi_{\text{th}}(j, i)$ est le seuil masquant à i Bark en raison de la composante masquante située à j Bark. $X(j)$ est le niveau de la pression sonore (en dB) des composantes masquantes avec la bande critique d'indice j . $O(j)$ est le paramètre offset du seuil.

$$O(j) = \begin{cases} -1.525 - 0.275j - 4.5 \text{ dB} & \text{pour composante tonale} \\ -1.525 - 0.175j - 0.5 \text{ dB} & \text{pour composante non tonale} \end{cases} \quad (\text{IV.9})$$

$SF(j, i)$ est la fonction d'étalement destinée pour le masquage inter bande, donnée par :

$$SF(j, i) = \begin{cases} 17(dz + 1) - 0.4X(j) - 6 \text{ dB} & -3 \leq dz < -1 \\ (-0.4X(j) + 6)dz \text{ dB} & -1 \leq dz < 0 \\ -17dz \text{ dB} & 0 \leq dz < 1 \\ -(dz - 1)(17 - 0.15X(j)) - 17 \text{ dB} & 1 \leq dz < 8 \end{cases} \quad (\text{IV.10})$$

L'estimateur de Blackman-Tukey est calculé, à partir de (IV.5) et (IV.6), comme suit [8] :

$$\Phi_B = \frac{1}{K} V \lambda \quad (\text{IV.11})$$

Φ_B est utilisé pour calculer le seuil Φ_{th} . Notons que le niveau de la pression sonore en Bark est donné par :

$$X(z(f_s w/W)) = \Phi_B(W), \quad w = 0, \dots, W - 1 \quad (\text{IV.12})$$

Avec w le nombre de points de la TFD, f_s la fréquence d'échantillonnage et $z(.)$ est donné par l'équation (IV.7). Cette information perceptuelle peut être représentée dans le domaine propre en utilisant l'équation (IV.5). Elle est donnée comme suit :

$$\theta = \frac{1}{W} V^T \Phi_{th} \quad (\text{IV.13})$$

4. **Cas du bruit coloré :** Application de IFET (Inverse Frequency Eigen-domain Transformation) pour le calcul de l'énergie du bruit ξ_i . Cette solution consiste à remplacer la variance du bruit σ_n^2 dans (III.49) avec l'énergie du bruit dans la direction des vecteurs propres de la matrice de covariance. Ceci peut être obtenu, en calculant le quotient de Railegh ξ_i , associé avec q_i et \hat{R}_{nn} pour $i = 1 \dots N$, comme suit :

$$\xi_i = q_i^T R_{nn} q_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_n(\omega) |V_i(\omega)|^2 d\omega \quad (\text{IV.14})$$

$$\xi = \frac{1}{W} V^T \Phi_n \quad (\text{IV.15})$$

5. **KLT :** Multiplication du vecteur signal par Q_1^T pour obtenir les coefficients du signal dans le sous-espace signal;
6. **Matrice du gain :** Multiplication des amplitudes par la matrice diagonale du gain G_2 donnée par :

$$G_2(i) = \exp(-v \xi_i / \min(\lambda_i, \theta_i)) \quad (\text{IV.16})$$

où v est un paramètre permettant de contrôler le compromis entre le niveau du bruit résiduel et la distorsion du signal.

IV.4.1. Description de la méthode "PKLT-VRE" proposée

La méthode PKLT-VRE proposée [51] est résumée dans le schéma bloc de la figure IV.3. Elle consiste à combiner deux méthodes de rehaussement en sous-espace, à savoir KLT-VRE et PKLT et à proposer, en deuxième lieu, une nouvelle matrice de gain, calculée en fonction des matrices G_1 et G_2 , que nous avons optimisée par une fonction de régression linéaire. Pour ce faire, nous avons procédé à la décomposition en valeurs propres (EVD) de la matrice d'autocorrélation \hat{R}_{xx} et au calcul du nombre optimal l de composantes principales, utilisées pour obtenir la matrice tronquée des valeurs propres $\Lambda(l)$. La matrice de gain proposée G (IV.17) est une fonction du maximum de G_1 et G_2 .

Les étapes d'implémentation de la méthode PKLT-VRE proposée sont comme suit :

1. **Matrice de covariance** : Calcul de la matrice de covariance \hat{R}_{xx} du signal bruité ;
2. **Décomposition en valeurs propres** : Décomposition de \hat{R}_{xx} en valeurs et vecteurs propres ;
3. **Utilisation du critère VRE** : Calcul du nombre optimal l de Composantes Principales (équation (III.120)), utilisé pour former $\Lambda(l)$
4. **Estimation du bruit** : Utilisation d'un Détecteur d'Activité Vocale (VAD) pour l'estimation et la mise à jour du bruit σ_n^2
5. **Calcul du gain** : A partir de $\Lambda(l)$ et σ_n^2 on déduit la valeur de G_1 (III. 123)
6. **Nouvelle matrice de gain** : Calcul à partir de (III.123) et (IV.16), de la nouvelle matrice de gain comme suit :

$$G = \text{Max}(G_1, G_2). \quad (\text{IV.17})$$

7. **KLT inverse** : Utilisation de la transformée IKLT (Q_1) (III.36) pour la reconstruction du signal rehaussé $\hat{s}(t)$.

IV.4.2. Optimisation de la matrice de gain

Nous avons également procédé à l'optimisation de la nouvelle matrice de gain par une fonction de régression linéaire [51], donnée comme suit :

$$G = \text{Max}(G_1, G_2) + a_1 \cdot G_1 + a_2 \cdot G_2, \quad a_1 \text{ and } a_2 \in \mathbb{R} \quad (\text{IV.18})$$

Le schéma bloc de la méthode PKLT-VRE proposée est donné à la figure IV.3

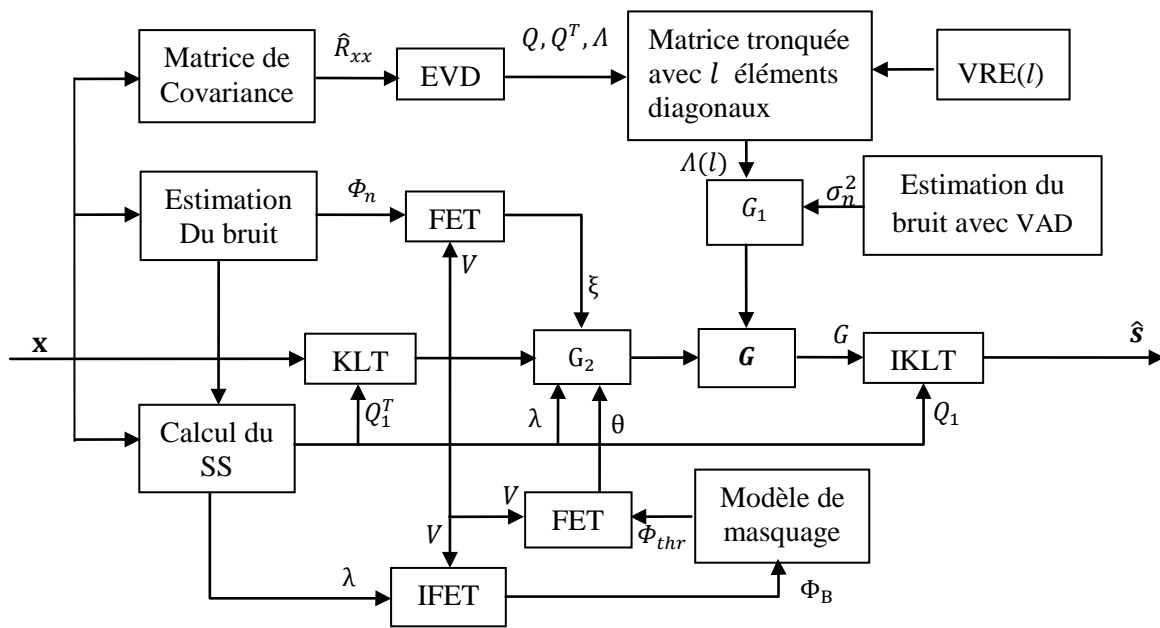


Figure IV.3 : Schéma bloc de la méthode PKLT-VRE proposée [51]

IV.4.3. Résultats et discussions

L'évaluation de la méthode PKLT-VRE proposée a été réalisée par SIG, BAK, OVL, SegSNR, WSS, PESQ et LLR [97]. Pour des considérations liées à la validité des résultats obtenus, nous avons opté pour deux bases de données disjointes NOIZEUS et TIMIT. Cette méthode a été comparée aux méthodes de Soustraction Spectrale (SS), de Wiener, de KLT-VRE et de PKLT [42]

Pour la méthode SS, plusieurs tests ont été conduits avec différentes valeurs du facteur d'échelle relatif à la période du silence F et du facteur de lissage utilisé dans la mise à jour du spectre du bruit γ . Les figures IV.4. (a1 à d1), montrent les spectrogrammes du signal bruité et rehaussé par la méthode SS, où nous pouvons remarquer (Figure IV.4. d1) que les meilleurs résultats sont obtenus pour $F = 0.003$ et $\gamma = 0.9$.

Pour le filtrage de Wiener, les tests ont été conduits pour différentes valeurs du facteur de lissage de la décision directe α et du facteur de lissage utilisé pour la mise à jour du spectre du bruit β . Les spectrogrammes du signal bruité et rehaussé relatif au filtrage de Wiener sont indiqués sur les figures IV.4. (a2 à d2). Les meilleurs résultats sont obtenus avec $\alpha = 0.99$ et $\beta = 0.99$ (Figure IV.4. d2).

Pour KLT-VRE, les tests ont été conduits pour différentes valeurs du multiplicateur de Lagrange μ_L et de $N = 21$. Les figures IV.4. (a3 à d3), montrent les spectrogrammes du signal bruité et rehaussé par la méthode KLT-VRE. Les meilleurs résultats sont obtenus avec $\mu_L = 2$ (figure IV.4. b3).

Pour PKLT, les tests ont été conduits pour différentes valeurs de ν (paramètre contrôlant le compromis entre le niveau du bruit résiduel et la distorsion du signal). Les figures IV.4. (a4 à d4), montrent les spectrogrammes du signal bruité et rehaussé par la méthode PKLT. Les meilleurs résultats sont obtenus avec $\nu = 0.05$ (figure IV.4. b4).

S'agissant de la méthode PKLT-VRE proposée, plusieurs tests ont été conduits pour différentes valeurs de ν et μ_L . Les figures IV.4. (a5 à d5), montrent les spectrogrammes du signal bruité et rehaussé par la méthode PKLT-VRE. Les meilleurs résultats sont obtenus avec $\nu = 0.05$ et $\mu_L = 2$ (figure IV.4. c5).

Il ressort de l'analyse des spectrogrammes que la méthode PKLT-VRE proposée présente les meilleurs résultats, comparée aux méthodes SS, Wiener, KLT-VRE et PKLT.

Les performances de la méthode proposée ont été confirmées par les résultats de l'évaluation subjective et objective, donnés sur les tableaux (IV.2, 3, 4, 5, 6 et 7).

Le tableau IV.2, présente les résultats de l'évaluation de la méthode PKLT-VRE proposée comparée avec SS, Wiener, PKLT et KLT-VRE en utilisant 20 phrases issues de la base de données NOIZEUS, corrompues par des bruits de chahut, Bus et rue. Les mesures LLR et WSS indiquent une perte de la parole et donc devraient être minimales. Les résultats obtenus montrent que SIG, BAK et OVRL sont généralement plus grands dans le cas de la méthode proposée.

L'évaluation de l'efficacité de cette méthode a été également effectuée en utilisant des bases de données disjointes. Nous avons choisi 10 phrases issues de la base de données TIMIT que nous avons bruité par trois types de bruit à savoir : buccaneer1_2, factory1_2 et babble_2, issues de la base de données NOISEX'92. Les résultats obtenus sont reportés sur les tableaux IV.3, IV. 4 et IV.5.

Nous remarquons, à partir du tableau IV.3, que les résultats obtenus sont, en majorité, nettement meilleurs que ceux des méthodes de Wiener et PKLT. Par exemple, pour la mesure PESQ, les différences enregistrées entre PKLT-VRE/ Wiener et PKLT-VRE / PKLT à 0dB de SNR sont respectivement 0.68 et 0.92. Pour SegSNR, les différences enregistrées entre PKLT-VRE/ Wiener et PKLT-VRE / PKLT à 0dB de RSB sont respectivement 1.78 et 1.97. nous pouvons également remarquer que la méthode proposée donne de meilleurs résultats avec la base de données TIMIT.

La figure IV.5, représente les résultats obtenus avec les mesures PESQ, SegSNR et WSS pour trois types de bruits à des niveaux de RSB de 0, 5 et 10 dB. Nous avons également reporté les résultats de l'évaluation objective relatifs aux données non traitées (signal bruité) utilisées dans ce cas comme données de référence. Les résultats obtenus montrent que la méthode proposée présente des valeurs élevées de SegSNR et PESQ et une valeur faible de WSS.

Nous avons reporté sur le tableau IV.6, les résultats de l'étude comparative entre G_1 et G_2 . Les résultats de l'optimisation de G (IV.19) indiquent une nette amélioration pour des valeurs $a_1 = 0.08$ et $a_2=1.2$.

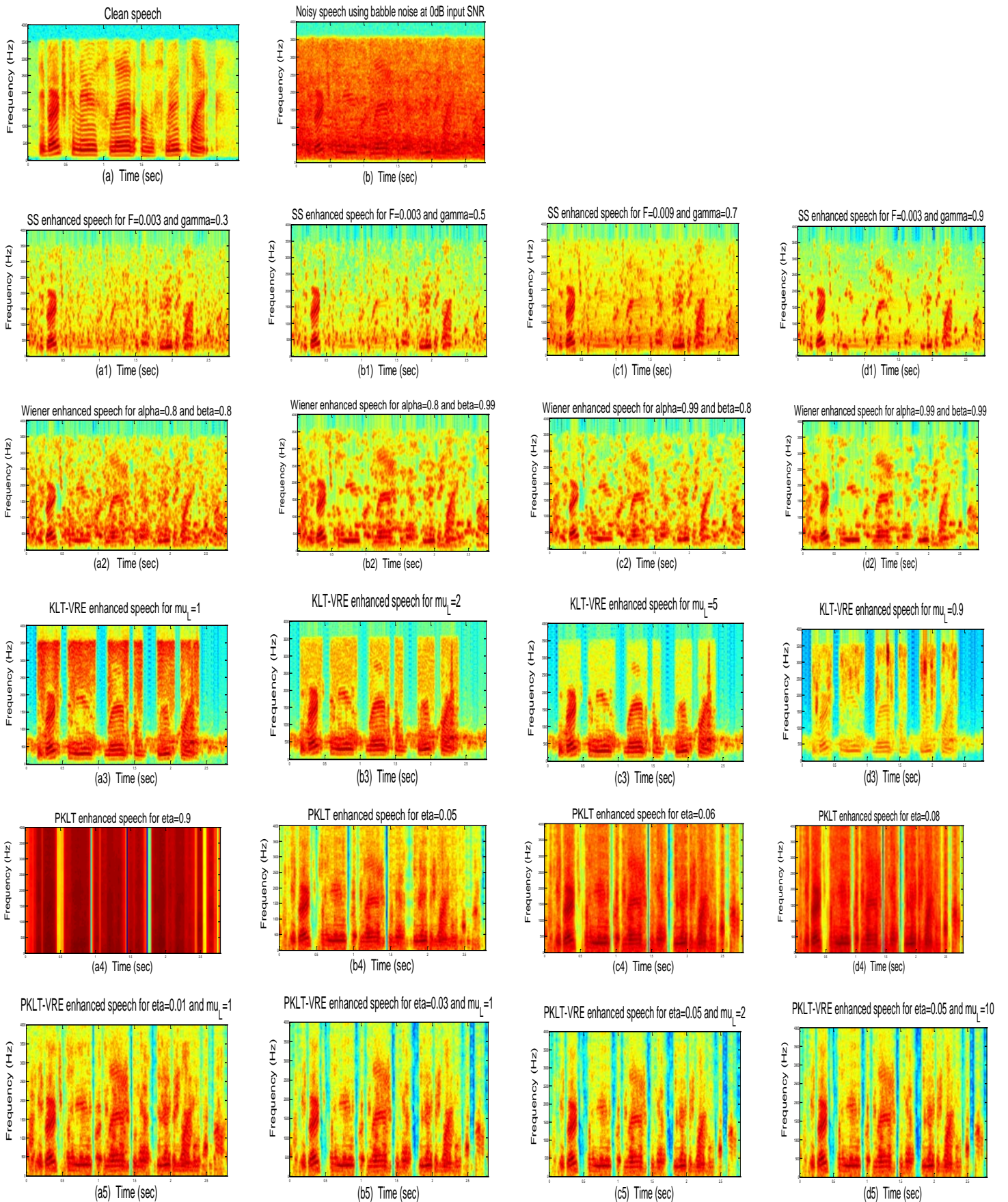


Figure IV.4 : Spectrogrammes des méthodes de rehaussement de la parole : (a) parole propre (phrase "the birch canoe slid on the smooth planks"), (b) parole bruité dégradée par un bruit de chahut à 0 dB de RSB d'entrée , (a1 à d1) Spectrogrammes de la méthode de SS pour différentes valeurs de F et γ ; (a2 à d2) Spectrogrammes de la méthode de Wiener pour différentes valeurs de α et β ; (a3 à d3) Spectrogrammes de la méthode KLT-VRE pour différentes valeurs de μ_L ; (a4 à d4) Spectrogrammes de la méthode PKLT pour différentes valeurs de μ_L et ν ; (a5 à d5) Spectrogrammes de la méthode proposée PKLT-VRE pour différentes valeurs de μ_L et ν .

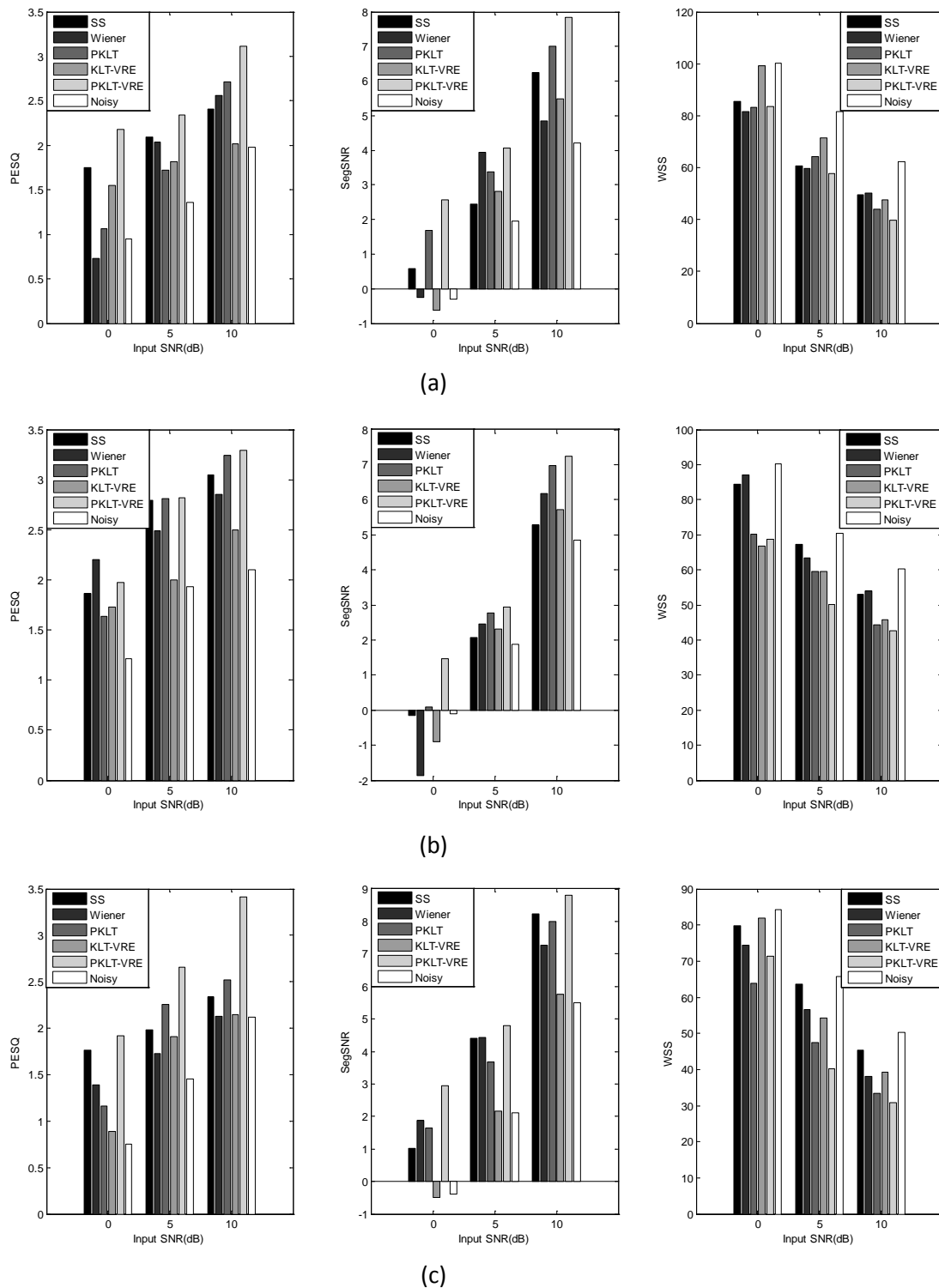


Figure IV.5 : Evaluations objectives avec PESQ, SegSNR et WSS pour différentes méthodes en utilisant : (a) bruit de "bus" avec NOIZEUS, (b) bruit de "rue" avec NOIZEUS, (c) bruit de "chahut" avec TIMIT.

Plusieurs tests de comparaison statistique basés sur l'analyse de la variance (p-value) ont été conduits pour évaluer les différences significatives entre la méthode proposée et les autres méthodes. Il est important de noter que même si cette approche n'est pas fréquente dans le

domaine du rehaussement de la parole, nous estimons qu'elle peut fournir un outil complémentaire pour évaluer l'efficacité de la méthode proposée.

Les résultats sont donnés dans le tableau IV.7 avec des mesures SIG, PESQ et SegSNR. Les différences entre les scores sont considérées comme statistiquement significatives si la valeur-P (niveau de signifiante) obtenue est inférieure à 0.05 (5%). En d'autres termes l'hypothèse nulle (méthode proposée atteint des performances égales ou inférieures à celles des autres méthodes) peut être rejetée si $p \leq 0.05$.

Nous pouvons montrer, à titre indicatif à partir du tableau IV.7, les résultats suivants :

- En termes de SIG, les tests sont statistiquement significatifs ($p \leq 0.05$), excepté pour les cas suivants : Wiener avec les bruits "buccaneer 1_2" et "factory 1_2", PKLT avec les bruits "babble_2" et SS avec le bruit "factory1_2", où $p > 0.05$;
- En termes de PESQ, les tests sont statistiquement significatifs ($p \leq 0.05$), excepté pour les cas suivants : Wiener avec les bruits "buccaneer 1_2" et "factory 1_2" et SS avec le bruit "buccaneer 1_2", où $p > 0.05$;
- En termes de SegSNR, les tests sont statistiquement significatifs ($p \leq 0.05$), excepté pour les cas suivants : Wiener avec les bruits "car" et "street", SS avec le bruit "babble_2" et PKLT avec le bruit "street", où $p > 0.05$.

Tableau IV.2 : Evaluation objective de la méthode PKLT-VRE proposée comparée avec SS, Wiener, PKLT et KLT-VRE et corrompue par des bruits de bus, rue et chahut. Valeurs moyenne obtenues en utilisant 20 phrases extraites de la base de données NOIZEUS. Les meilleures performances sont indiquées en Gras.

Objective Measures	Input SNR (dB)	Car noise						Street noise				
		No process	SS	Wiener	PKLT	KLT-VRE	PKLT-VRE	SS	Wiener	PKLT	KLT-VRE	PKLT-VRE
SIG [1 to 5]	0	1.27	2.27	1.35	1.63	1.15	2.34	2.24	2.23	1.99	2.24	2.75
	5	1.67	2.74	2.30	2.83	1.87	2.68	2.45	2.39	2.61	2.43	2.82
	10	2.17	3.18	2.90	3.34	2.24	3.61	2.62	2.77	2.66	2.53	3.45
BAK [1 to 5]	0	1.22	2.20	2.08	1.66	1.65	1.45	2.10	1.86	1.85	2.21	2.11
	5	1.73	2.13	1.91	2.18	2.21	2.50	2.27	1.92	2.32	2.46	2.29
	10	2.03	2.45	2.37	2.81	2.32	2.81	2.87	2.26	2.50	2.54	2.55
OVRL [1 to 5]	0	0.94	1.92	0.85	1.12	1.15	2.08	2.02	2.15	1.98	2.17	2.33
	5	1.31	2.03	1.58	2.10	1.83	2.35	2.15	2.62	2.19	2.28	2.29
	10	2.11	2.38	3.05	2.19	1.95	2.81	2.23	2.88	2.70	2.34	2.99
PESQ [0.5- 4.5]	0	0.95	1.75	0.73	1.06	1.55	2.18	1.86	2.20	1.64	1.73	1.97
	5	1.36	2.09	2.04	1.72	1.82	2.34	2.79	2.49	2.81	2.00	2.82
	10	1.98	2.41	2.56	2.71	2.02	3.11	3.05	2.85	3.24	2.50	3.29
SegSNR	0	-0.30	0.59	-0.25	1.68	-0.63	2.58	-0.15	-1.86	0.08	-0.90	1.46
	5	1.95	2.45	3.93	3.37	2.82	4.06	2.08	2.46	2.78	2.31	2.95
	10	4.20	6.25	4.84	7.00	5.48	7.84	5.28	6.18	6.97	5.71	7.24
WSS	0	100.31	85.43	81.46	83.25	99.34	83.46	84.53	87.08	70.08	66.68	68.83
	5	81.42	60.77	59.75	64.17	71.53	57.81	67.33	63.43	59.51	59.44	50.09
	10	62.30	49.65	50.27	44.02	47.54	39.58	52.93	53.94	44.38	45.76	42.66
LLR	0	1.94	1.38	0.77	1.12	1.89	0.96	0.99	0.98	1.28	1.40	1.05
	5	1.34	0.87	0.54	0.82	1.51	0.75	0.87	0.78	1.09	1.03	0.67
	10	1.02	0.63	0.48	0.65	1.32	0.47	0.81	0.74	0.97	1.09	0.39

Objective Measures	Input SNR (dB)	Babble noise				
		SS	Wiener	PKLT	KLT-VRE	PKLT-VRE
SIG [1 to 5]	0	1.31	1.78	1.61	1.41	2.18
	5	1.72	1.93	2.28	1.99	2.85
	10	2.18	2.56	2.94	2.48	3.12
BAK [1 to 5]	0	1.42	1.74	1.63	1.54	1.83
	5	1.97	2.23	2.17	2.18	2.76
	10	2.07	2.71	2.62	2.47	2.94
OVRL [1 to 5]	0	1.20	1.83	1.58	1.33	1.94
	5	1.68	1.93	1.82	2.17	2.45
	10	2.02	2.38	2.35	2.24	2.63
PESQ [0.5 to 4.5]	0	1.21	2.41	1.32	2.04	2.45
	5	1.54	2.67	1.71	2.15	2.93
	10	1.76	3.27	2.07	2.35	2.86
SegSNR	0	-4.48	-3.22	0.63	-4.65	-2.64
	5	-0.12	1.47	2.34	0.42	2.47
	10	2.30	3.18	3.28	2.49	4.09
WSS	0	92.23	87.35	77.30	81.06	71.34
	5	85.09	60.99	62.09	64.84	67.84
	10	76.56	48.37	45.65	54.61	50.21
LLR	0	2.43	1.53	1.82	2.11	1.24
	5	1.89	1.08	1.25	1.76	0.94
	10	1.14	0.82	0.95	1.49	0.71

Tableau IV.3 : Evaluation objective en utilisant des données disjointes composées de 10 phrases extraites des bases de données NOIZEUS et TIMIT. La méthode PKLT-VRE proposée est comparée avec SS, Wiener, PKLT et KLT-VRE et corrompue avec le bruit de chahut. Les meilleures performances sont indiquées en Gras.

Objective Measures	Input SNR (dB)	Evaluation tests using NOIZEUS database					Evaluation tests using TIMIT database				
		SS	Wiener	PKLT	KLT-VRE	PKLT-VRE	SS	Wiener	PKLT	KLT-VRE	PKLT-VRE
SIG [1 to 5]	0	2.06	2.16	2.53	1.56	2.98	2.02	2.43	1.75	1.34	2.41
	5	2.28	2.08	3.09	2.18	3.14	2.30	2.61	2.62	1.38	2.89
	10	2.81	2.79	3.16	2.82	3.48	2.39	2.96	3.11	1.76	3.66
BAK [1 to 5]	0	2.08	2.22	2.15	1.67	2.34	1.96	1.78	1.76	1.70	1.91
	5	2.75	2.45	2.52	2.03	2.66	1.74	1.90	2.00	1.67	2.93
	10	2.95	2.50	3.00	2.18	3.87	2.09	2.27	2.14	2.36	3.56
OVL [1 to 5]	0	2.15	2.46	2.08	1.44	2.37	1.93	1.76	2.21	1.69	2.25
	5	2.39	2.52	2.49	1.97	2.95	2.19	2.44	2.49	1.93	2.59
	10	2.97	2.49	3.11	2.02	3.54	2.40	2.84	2.55	2.26	2.98
PESQ [0.5 to 4.5]	0	1.89	1.54	1.97	1.62	2.12	1.76	1.39	1.16	0.89	1.92
	5	2.12	1.93	2.05	1.98	2.97	1.98	1.73	2.26	1.91	2.66
	10	2.68	2.46	2.25	2.14	3.04	2.34	2.13	2.52	2.15	3.41
SegSNR	0	1.21	1.52	1.33	-0.18	3.30	1.02	1.87	1.64	-0.5	2.94
	5	4.71	4.78	3.90	2.12	4.95	4.41	4.43	3.67	2.16	4.80
	10	8.07	8.24	7.62	5.26	9.13	8.23	7.28	8.01	5.76	8.80
WSS	0	83.44	81.06	66.09	75.09	64.87	79.85	74.48	63.98	81.86	71.42
	5	61.84	57.63	47.84	67.94	50.68	63.75	56.49	47.52	54.32	40.16
	10	52.70	45.68	37.70	56.41	37.60	45.36	37.98	33.37	39.15	30.82
LLR	0	0.72	0.62	1.00	1.72	0.82	0.97	0.94	1.24	1.73	1.02
	5	0.66	0.49	0.82	1.45	0.44	0.81	0.56	0.87	1.62	0.49
	10	0.48	0.42	0.50	1.32	0.37	0.50	0.49	0.60	1.18	0.32

Tableau IV.4 : Evaluation objective en utilisant des données disjointes composées de 10 phrases extraites des bases de données NOIZEUS et TIMIT. La méthode PKLT-VRE proposée est comparée avec SS, Wiener, PKLT et KLT-VRE et corrompue avec le bruit d'avion de chasse. Les meilleures performances sont indiquées en Gras.

Objective Measures	Input SNR (dB)	Evaluation tests using NOIZEUS database					Evaluation tests using TIMIT database				
		SS	Wiener	PKLT	KLT-VRE	PKLT-VRE	SS	Wiener	PKLT	KLT-VRE	PKLT-VRE
SIG [1 to 5]	0	1.95	1.54	1.97	1.83	2.34	1.76	1.66	2.23	1.63	2.43
	5	2.19	2.50	2.32	2.10	2.77	1.97	2.37	2.56	2.09	2.95
	10	2.50	2.97	2.83	2.24	3.12	2.16	2.74	2.68	2.27	3.31
BAK [1 to 5]	0	1.70	1.39	1.97	1.69	2.22	1.75	1.87	1.89	1.75	1.91
	5	2.26	2.09	2.22	1.72	2.64	1.80	2.06	1.93	1.80	2.07
	10	2.57	2.20	2.48	2.17	2.81	2.04	2.20	2.11	1.97	2.17
OVL [1 to 5]	0	2.06	1.39	1.77	1.60	2.12	1.99	1.60	1.73	1.56	2.15
	5	2.29	2.21	2.01	1.73	2.57	2.20	2.17	2.14	1.93	2.45
	10	2.36	2.24	2.28	1.94	3.14	2.13	2.21	2.51	2.10	3.74
PESQ [0.5 to 4.5]	0	1.24	1.97	1.71	1.45	2.15	1.17	1.33	1.00	0.98	1.97
	5	2.23	2.45	1.74	2.35	2.83	2.11	2.47	1.54	1.55	2.42
	10	2.30	3.50	2.75	2.61	3.67	2.46	3.06	2.77	2.10	3.67
SegSNR	0	1.14	0.97	2.16	-0.61	1.75	0.86	1.21	2.43	0.49	2.38
	5	4.09	4.13	4.90	2.75	4.85	3.90	5.14	4.92	2.69	5.82
	10	6.99	7.20	7.95	4.81	8.24	6.58	7.81	8.20	4.94	8.81
WSS	0	95.94	81.48	84.77	91.84	83.22	90.66	85.29	80.15	83.89	80.01
	5	75.83	79.71	59.13	65.29	58.03	68.33	58.71	56.72	62.48	59.37
	10	65.54	53.89	52.67	57.13	50.23	48.29	55.19	37.78	42.67	37.51
LLR	0	1.64	1.40	1.69	2.22	1.57	1.29	1.21	1.25	1.71	1.20
	5	1.13	0.95	1.28	1.73	0.91	0.83	1.03	0.97	1.49	0.88
	10	0.81	0.64	0.96	1.67	0.80	0.65	0.88	0.63	1.25	0.57

Tableau IV.5 : Evaluation objective en utilisant des données disjointes composées de 10 phrases extraites des bases de données NOIZEUS et TIMIT. La méthode PKLT-VRE proposée est comparée avec SS, Wiener, PKLT et KLT-VRE et corrompue avec le bruit d'usine. Les meilleures performances sont indiquées en Gras.

Objective Measures	Input SNR (dB)	Evaluation tests using NOIZEUS database					Evaluation tests using TIMIT database				
		SS	Wiener	PKLT	KLT-VRE	PKLT-VRE	SS	Wiener	PKLT	KLT-VRE	PKLT-VRE
SIG [1 to 5]	0	1.46	1.67	1.31	1.16	1.40	1.55	1.89	1.24	1.26	1.70
	5	1.97	2.19	1.83	1.47	2.74	1.98	2.13	1.95	1.76	2.69
	10	2.59	2.71	2.31	1.75	3.72	2.65	2.59	2.43	2.08	3.04
BAK [1 to 5]	0	1.40	1.58	1.56	1.51	1.69	1.18	1.29	1.22	0.74	1.92
	5	1.88	1.92	1.94	1.84	1.97	1.52	1.53	1.54	1.03	2.37
	10	2.27	2.34	2.27	2.16	2.30	2.11	2.01	1.96	1.39	2.86
OVRL [1 to 5]	0	1.73	1.75	1.51	1.20	1.80	1.67	1.56	1.52	1.05	1.82
	5	2.12	2.15	2.08	1.93	2.24	2.10	1.86	2.09	1.76	2.38
	10	2.43	2.39	2.74	2.58	2.91	2.97	2.77	2.49	2.18	3.24
PESQ [0.5 to 4.5]	0	1.24	1.83	1.87	1.03	1.92	1.40	1.76	1.81	0.95	1.97
	5	2.34	2.47	2.48	1.84	2.33	1.92	2.19	2.36	1.46	2.53
	10	2.49	2.66	2.57	2.13	2.99	2.25	2.58	2.47	2.02	3.18
SegSNR	0	1.33	2.31	1.99	0.91	2.18	1.16	1.52	1.40	-1.58	2.77
	5	4.09	5.14	4.85	3.77	6.31	3.85	3.65	3.90	1.96	3.71
	10	7.92	8.14	8.15	6.86	9.43	7.36	7.65	6.24	3.45	8.82
WSS	0	94.92	77.37	78.47	77.75	72.95	101.2	99.02	98.88	118.69	100.42
	5	71.23	52.53	54.25	55.21	53.30	92.67	83.23	96.72	117.12	94.11
	10	47.95	37.79	36.20	35.63	32.56	59.31	68.37	66.69	82.06	55.24
LLR	0	1.18	1.12	1.47	1.64	1.40	1.31	1.14	1.34	1.84	1.06
	5	0.90	1.15	1.22	1.31	0.83	0.62	0.52	0.76	1.59	0.72
	10	0.62	0.67	0.90	1.02	0.59	0.43	0.35	0.55	1.41	0.51

Tableau IV.6 : Evaluation objective de la méthode PKLT-VRE avec optimisation du gain. Cette comparaison est utilisée avec 3 phrases issues de la base de données TIMIT.

Objective Measures	Input SNR (dB)	$G = \text{Max}(G_1, G_2)$	$G = \text{Max}(G_1, G_2) + a_1 G_1 + a_2 G_2$
		PKLT-VRE	PKLT-VRE
SIG [1 to 5]	0	1.49	1.87
	5	2.14	2.88
	10	2.53	3.02
SIG [1 to 5]	0	1.74	1.93
	5	2.24	2.94
	10	2.43	3.15
OVRL [1 to 5]	0	1.23	2.02
	5	1.81	2.60
	10	2.90	3.20
PESQ [0.5 to 4.5]	0	1.29	1.67
	5	2.12	2.28
	10	2.89	3.10
SegSNR	0	1.62	2.15
	5	4.30	4.39
	10	7.34	8.06
WSS	0	87.15	86.33
	5	67.15	62.34
	10	49.22	40.18
LLR	0	1.55	1.08
	5	1.32	0.93
	10	0.94	0.57

Tableau IV.7 : les valeurs-P ($p < 0.05$) des mesures objectives SIG, PESQ et SegSNR de la technique PKLT-VRE comparée à SS, Wiener, PKLT et KLT_VRE pour différents types de bruits à 0 dB de RSB d'entrée.

Type of noise	Objectives evaluation	P-value of the subspace speech enhancement methods			
		PKLT-VRE/ SS	PKLT-VRE / wiener	PKLT-VRE/ PKLT	PKLT-VRE / KLT-VRE
Street NOIZEUS database	SIG	0.026	0.008	0.044	0.031
Buccaneer jet NOIZEUS database		0.008	>0.05	0.007	0.011
Babble NOIZEUS database		0.004	0.007	>0.05	0.022
Buccaneer jet TIMIT database		0.010	0.005	0.041	0.003
Factory TIMIT database		>0.05	>0.05	0.008	0.022
Car NOIZEUS database	PESQ	0.036	>0.05	0.039	0.025
Factory TIMIT database		0.010	>0.05	0.009	0.007
Buccaneer jet TIMIT database		>0.05	>0.05	0.015	0.020
Car NOIZEUS database	SegSNR	0.002	>0.05	0.002	0.028
Babble TIMIT database		>0.05	0.040	0.005	0.030
Street NOIZEUS database		0.022	>0.05	>0.05	0.046

IV.5. Conclusion

Dans ce chapitre, nous avons présenté deux nouvelles méthodes de rehaussement de la parole par décomposition en sous-espaces.

La première méthode a concerné l'introduction du critère VRE et de la méthode MCRA en vue d'optimiser le modèle KLT de référence et d'estimer de façon continue la DSP du bruit.

La deuxième méthode a consisté en l'introduction du critère VRE dans le modèle PKLT de référence en vue d'optimiser les paramètres de ce dernier suivi de la proposition d'une nouvelle matrice de gain de suppression que nous avons optimisée par une fonction de régression linéaire.

La robustesse de ces deux méthodes a été confirmée par les résultats de l'évaluation objective, reportés sur les différents tableaux et spectrogrammes en rapport avec les signaux propres, bruités et rehaussés.

Conclusions et perspectives

1. Conclusions

Dans cette thèse, nous avons abordé le problème du rehaussement de la parole dégradée par un bruit additif prélevé à partir de la base de données Noisex [106]. Le corpus de parole que nous avons utilisé est issu des bases de données NOIZEUS et TIMIT [107]. Le choix de ces deux bases de données est motivé par le fait que NOIZEUS nous a permis, d'une part, l'utilisation des mêmes phrases phonétiquement équilibrées, des mêmes types de bruits et des mêmes mesures objectives qui ont été déjà utilisés pour évaluer les méthodes de référence (SS, Wiener, LSA, KLT-VRE, PKLT). D'autre part, nous avons également jugé utile de disposer d'une base de données disjointe TIMIT, qui nous a permis de confirmer la robustesse des méthodes proposées par la mesure des degrés de significations statistiques des résultats obtenus par les deux bases de données.

Comme nous l'avons déjà évoqué, les algorithmes de rehaussement de la parole ne parviennent généralement pas à supprimer le bruit additif et maintenir, en même temps, l'intégralité du signal propre. Cette opération de rehaussement est donc souvent réalisée en défaveur de l'intelligibilité du signal rehaussé.

Idéalement, ces algorithmes doivent être en mesure d'améliorer à la fois la qualité et l'intelligibilité du signal. Cependant, en pratique ces algorithmes agissent uniquement sur la qualité par le biais de la réduction du bruit additif tout en introduisant des distorsions affectant plus ou moins l'intelligibilité du signal reçu désignée par bruit résiduel ou musical.

Aussi, l'estimation du bruit peut avoir un impact majeur sur la qualité du signal rehaussé. Si l'estimation du bruit est trop faible, le bruit résiduel devient audible et si elle est trop élevée, la parole sera distordue entraînant ainsi une perte d'intelligibilité.

La méthode la plus simple consiste à estimer et mettre à jour le spectre du bruit pendant les segments de silences du signal en utilisant un algorithme de détection d'activité vocale (VAD). Même si une telle approche peut fonctionner de manière satisfaisante dans le cas d'un

bruit stationnaire, elle demeure cependant inadaptée pour des environnements plus réalistes, où les caractéristiques spectrales du bruit changent constamment.

C'est dans cette optique que s'est inscrit notre travail de recherche qui consiste, premièrement, à développer des méthodes hybrides basées sur la décomposition en sous-espaces de la parole bruitée en tenant en compte des aspects liés au nombre optimal des Composantes Principales devant être utilisées pour reconstruire le signal propre, à l'estimation continue de la DSP du bruit non stationnaire et au phénomène de masquage du bruit qui consiste à réduire l'effet perceptuel du bruit musical.

Nous avons présenté en chapitre II les méthodes de rehaussement de la parole de l'état de l'art, basées sur le principe de la Soustraction Spectrale, du filtrage de Wiener, du filtrage de Kalman et de la théorie de l'estimation statistique avant d'entamer en chapitre III, les méthodes de décomposition en sous-espaces. Dans ce cadre, nous avons introduit le concept de l'Analyse en Composante Principales (PCA) ou KLT et le critère de sélection VRE qui permet une meilleure reconstruction des variables par la sélection du nombre optimal de Composantes Principales (CP). Il a été également question, dans cette thèse, de considérer le cas du bruit non stationnaire que nous avons estimé sa DSP à partir de la méthode (MCRA).

Dans ce travail, nous avons proposé deux nouvelles méthodes de rehaussement de la parole. La première méthode proposée était basée principalement sur l'intégration avec la méthode de référence le critère VRE et la méthode MCRA destinés respectivement pour l'optimisation du modèle KLT et l'estimation de la DSP du bruit. La deuxième méthode proposée a concerné l'optimisation des paramètres KLT par le critère VRE et la proposition d'une nouvelle matrice de gain que nous avons optimisé par une fonction de régression linéaire.

L'évaluation de ces deux méthodes a été menée par un ensemble de mesures objectives dont les résultats obtenus, confirmés par les spectrogrammes des signaux propres, bruités et rehaussés, ont été très satisfaisants.

2. Perspectives

Les perspectives ouvertes à ce travail, concernent principalement les axes de recherche suivants :

- Amélioration des performances de la méthode PKLT-VRE proposée, à travers l'estimation de la DSP du bruit par la technique MCRA ;
- Les performances de reconnaissance de la parole dans des environnements acoustiques réels pourraient être également améliorées par l'introduction de cette approche, à l'effet de lutter contre le problème de non-stationnarité du bruit, largement évoqué dans les Systèmes de Reconnaissance Automatique de la Parole (RAP) ou du locuteur (RAL).

Références bibliographiques

- [1] P. C. Loizou, *Speech Enhancement Theory And Practice*, 1st Edition, CRC Press (2007).
- [2] A. Acero, *Acoustical And Environmental Robustness In Automatic Speech Recognition*, Kluwer academic Publisher, 1993.
- [3] Md. F.R. Chowdhury, *A Soft Computing Approach For On-Line Automatic Speech Recognition In Highly Non-Stationary Acoustic Environments*, Ph.D in Telecommunications, INRS, université du Québec, 2012.
- [4] J.C.Junqua, J.Haton, *Robustness In Automatic Speech Recognition: Fundamentals End Apllications*, Kluwer Academic Publisher (1996).
- [5] M. Akbacak, J.H.L.Hansen, *Environmental Sniffing : Noise Knowledge Estimation For Robust Speech Systems*, *IEEE Trans. on ASLP*. 15(2) (2007) 465-477.
- [6] K. Pearsons, R. Bennett, S. Fidell, *Speech Levels In Various Noise Environments*, Technical Report EPA-600/1-77-025, US, Environmental Protection Agency.
- [7] S. Gazor, W. Zhang, *A Soft Voice Activity Detector Based On A Laplacian-Gaussian Model*, *IEEE Trans. on SAP*. 11(5) (2003) 498–505.
- [8] J. Ramírez, J. M. Górriz, J. C. Segura, *Voice Activity Detection. Fundamentals and Speech Recognition System Robustness*, ISBN 987-3-90213-08-0, (2007) 460, university of Granada Spain.
- [9] R. Martin, *Noise Power Spectral Density Estimation Based On Optimal Smoothing And Minimum Statistics*, *IEEE Trans. on SAP*. 9(5) (2001) 504-512.
- [10] D. Malah, R. V. Cox, A. J. Accardi, *Tracking Speech-Presence Uncertainty To Improve Speech Enhancement In Nonstationary Noise Environments*, In Proc. IEEE. ICASSP. (1999) 789–792.
- [11] I. Cohen , B. Berdugo, *Speech Enhancement For Non-Stationary Noise Environments*, Elsevier, *Signal Processing* 81 (2001) 2403–2418.
- [12] V.A. Martin, P. Pollack, *Methods For Speech SNR Estimation: Evaluation Tool And Analysis Of VAD Dependency*, *Radioengineering* . 14(1) (2005) 6-11.
- [13] I. Cohen, B. Berdugo, *Noise Estimation By Minima Controlled Recursive Averaging For Robust Speech Enhancement*, *IEEE SPL*. 9 (1) (2002) 12-15.
- [14] S. Rangachari, P.C. Loizou, *A Noise Estimation Algorithm For Highly Non-Stationary Environments*, *Speech Communication*, Elsevier, 28(2006) 220-231.
- [15] I. Cohen, *Noise Spectrum Estimation In Adverse Environments: Improved Minima Controlled Recursive Averaging*, *IEEE Trans. on SAP*. 11(5) (2003) 466-475.
- [16] G. Doblinger, *Computationally Efficient Speech Enhancement By Spectral Minima Tracking In Sub-bands*, in Proc. EUROSPEECH, 2 (1995) 1513–1516.
- [17] J. Sohn, W. Sung, *A Voice Activity Detector Employing Soft Decision Based Noise Spectrum Adaptation*, In Proc. IEEE. ICASSP. 1(1998) 365–368.
- [18] Berouti, R. Schwartz, J. Makhoul, *Enhancement Of Speech Corrupted By Acoustic Noise*, In Proc. IEEE. ICASSP. (1979) 208-211.
- [19] S. Boll, *Suppression Of Acoustic Noise In Speech Using Spectral Subtraction*, *IEEE Trans. on ASSP*. 27(2) (1979) 113-120.
- [20] J. S. Lim, A. V. Oppenheim, *Enhancement And Bandwidth Compression Of Noisy Speech*, In Proc of the IEEE. 67(12) (1979) 1586-1604.
- [21] R. J. McAulay, M. L. Malpass, *Speech Enhancement Using A Soft Decision Noise Suppression Filter*, *IEEE Trans. on ASSP*. 28(2) (1980) 137-145.
- [22] Y. Ephraim, D. Malah, *Speech Enhancement Using MMSE Short-Time Spectral Amplitude Estimator*, *IEEE Trans. on ASSP*. 2(2) (1984).
- [23] Y. Ephraim, D.Malah, *Speech Enhancement Using A Minimum Mean-Square Error Log-Spectral Amplitude Estimator*, *IEEE Trans. on ASSP*. 23(2) (1985) 443-445.
- [24] J.S. Erkelens, R.C. Hendriks, R. Heusdens, J.Jensen, *Minimum Mean-Square Error Estimation of Discrete Fourier Coefficients with Generalized Gamma Priors*, *IEEE Trans. on ASLP*. 15(6) (2007) 1741-1752.

- [25] P. J. Wolf, S. J. Godsill, Efficient Alternatives To The Ephraim And Malah Suppression Rule For Audio Signal Enhancement, *EURASIP J. Appl. Signal Process.* (10) (2003) 1043–1051.
- [26] R. Martin, Speech Enhancement Using A MMSE Short Time Spectral Estimation With Gamma Distributed Speech Priors, In *Proc. IEEE. ICASSP.* (2002) 253-256.
- [27] R. Martin, Speech Enhancement Base On MMSE And Super-gaussian Priors, *IEEE. Trans. on SAP.* 13(5) 845-856.
- [28] I. Cohen, Speech Enhancement Using Super-Gaussian Speech Models And Noncausal A Priori SNR Estimation, *Speech comm.* 47 (2005) 336-350.
- [29] I. Cohen, B. Berdugo, *Speech Enhancement For Non-Stationary Noise Environments*, Elsevier, *Signal processing* 81 (2001) 2403-2418.
- [30] K. R. Matthews, *Elementary linear algebra*, Department of Mathematics, University Of Queensland (2010).
- [31] G. Strang, *Linear Algebra And Its Applications*, 3rd ed. New York : Harcourt Brace Jovanovich, 1988.
- [32] Y. Ephraim, H. L. Van Trees, A Signal Subspace Approach For Speech Enhancement, *IEEE. Trans. on SAP.* 3(4) (1995) 251-266.
- [33] K. Hermus, P. Wambacq, H. V. hamme, A Review Of Signal Subspace Speech Enhancement And Its Application To Noise Robust Speech Recognition, *EURASIP J. Advances in Signal Processing*, Article ID 45821, 15 pages (2007)
- [34] U. Mittal, N. Phamdo, Signal/Noise KLT Based Approach For Enhancing Speech Degraded By Colored Noise, *IEEE Trans. SAP.* 8(2) (2000) 159–167.
- [35] A. Rezayee, S. Gazor, An Adaptive KLT Approach For Speech Enhancement, *IEEE Trans. SAP.* 9(2) (2001) 87–95.
- [36] H. Lev-Ari, Y. Ephraim, Extension Of The Signal Subspace Speech Enhancement Approach To Colored Noise, *IEEE SPL.* 10(4) (2003) 104-106.
- [37] Y. Hu, P. C. Loizou, A Generalized Subspace Approach For Enhancing Speech Corrupted By Colored Noise, *IEEE Trans. SAP.* 11(2003) 334–341.
- [38] J. Chen, J. Benesty, Y. Huang, Study Of The Noise-Reduction Problem In The Karhunen–Loève Expansion Domain, *IEEE Trans. on ASLP.* 17(4) (2009) 787-802.
- [39] B. Moor, The Singular Value Decomposition And Long And Short Spaces Of Noisy Matrices, *IEEE Trans. SP.* 41(9) (1993) 2826-2838.
- [40] M. Dendrinou, S. Bakamides, and G. Carayannis, *Speech Enhancement From Noise : A Regenerative Approach*, *Speech Communication*, Elsevier, 10 (1991) 45-57.
- [41] J. D. Johnston, Transform Coding Of Audio Signal Using Perceptual Noise Criteria, *IEEE J. Sel. Areas Comm.* 6(2) (1988) 314–323.
- [42] F. Jabloun, B. Champagne, Incorporating The Human Hearing Properties In The Signal Subspace Approach For Speech Enhancement, *IEEE Trans. on ASSP.* 11(6) (2003) 700-708.
- [43] C. H. You, S. N. Koh, S. Rahardja, An Invertible Frequency Eigen-Domain Transformation For Masking-Based Subspace Speech Enhancement, *IEEE S PL.* 12(5) (2005) 461-464.
- [44] J. U. Kim, S.G. Kim, C. D. Yoo, The Incorporation Of Masking Threshold To Subspace Speech Enhancement, In *Proc. IEEE. ICASSP.* 1 (2003) 76-79.
- [45] Y. Hu, P. C. Loizou, Incorporating A Psychoacoustical Model In Frequency Domain Speech Enhancement, *IEEE SPL.* 11(2) (2004) 270-273.
- [46] Y. Hu, P. C. Loizou, A Perceptually Motivated Approach For Speech Enhancement, *IEEE Trans. on ASSP.* 11(5) (2003) 457-465.
- [47] A. Ben Aicha, S. Ben Jebara, Perceptual Musical Noise Reduction Using Critical Band Tonality Coefficients And Masking Thresholds, In *Proc. Inter speech Conf*, Antwerp, (2007) 822-825.
- [48] A.H. Abolhassani, *Speech Enhancement Using Principal Component Analysis And Variance Of The Reconstruction Error Model Identification*, A Thesis Presented For The Master Degree Of Telecommunications, INRS-Telecommunications, University of Québec, 2007.
- [49] I. Ching-Ta, Enhancement Of Single Channel Speech Using Perceptual-Decision-Directed Approach, *Speech Comm. Elsevier*, 53(4) (2011) 495-507.

- [50] A. Saadoune, A. Amrouche, S.A. Selouani, MCRA Noise Estimation For KLT-VRE Based Speech Enhancement, *IJST. Springer Journals*, 16 (3) (2013) 333-339.
- [51] A. Saadoune, A. Amrouche, S.A. Selouani Perceptual Subspace Speech Enhancement Using Variance Of The Reconstruction Error, *DSP. Elsevier*, 24 (2014) 187–196.
- [52] J.O.Sopholes, SVD, PCA, KLT, CCA and All That, Rutgers University Electrical & Computer Engineering Department, OSP. (2002-2007) 332-525.
- [53] J. Gerbrands, On the relationships between SVD, KLT, and PCA, Department of Electrical Engineering, Delft University of Technology (Netherlands), *Pattern Recognition* , 14(6) (1981) 375-381.
- [54] A. Abolhassani, S.A. Selouani, D. O’Shaughnessy, M. F. Harkat, Speech Enhancement Using PCA And Variance Of The Reconstruction Error Model Identification, In *Proc. Inter speech. Belgium* (2007) 974-977.
- [55] T. FILLON, Traitement Numérique Du Signal Acoustique pour une Aide aux Malentendants, Thèse, École Nationale Supérieure des Télécommunications (2004).
- [56] O. Cappé, Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor, *IEEE Trans. SAP*. 2(2) (1994) 345-349.
- [57] T.,Hasan, M. K. Hasan, Suppression Of Residual Noise From Speech Signals using Empirical Mode Decomposition, *IEEE S. P L*. 16(1) (2009) 2-5.
- [58] F. Chen, P. C. Loizou, Impact Of SNR And Gain-Function Over- And Under-Estimation On Speech Intelligibility, *Elsevier, Speech Comm.* 54 (2012) 272–281.
- [59] P. Scalart, J. V. Filho, Speech Enhancement Based On A Priori Signal To Noise Estimation, In *Proc. IEEE. ICASSP 2* (1996) 629-632.
- [60] Y. Hu, P.C. Loizou, NOIZEUS: A Noisy Speech Corpus For Evaluation Of Speech Enhancement Algorithms. available at: <http://www.utdallas.edu/~loizou/speech/noizeus/>
- [61] P. Lockwood, J. Boudy, Experiments With A Non-Linear Speech Subtractor (NSS), Hidden Markov Models And Projections For Robust Recognition In Cars, *Speech Comm.* 11(2-3) (1992) 215-228.
- [62] S. Kamaths, P. Loizou, A Multi-Band Spectral Subtraction Method For Enhancing Speech Corrupted By Color Noise, In *Proc. IEEE. IC ASSP* (2002).
- [63] B. Sim, Y. Tong, J. Chang, A Parametric Formulation Of The Generalized Spectral Subtraction Method, *IEEE. Trans. SAP*. 6(4) (1998) 328-337.
- [64] H. Gustafsson, S. Nordholm, I. Claesson, Spectral subtraction using reduced delay convolution and adaptive averaging, *IEEE Trans. Speech Audio Process.*, 9(8) (2001)799-807. AGA
- [65] Akaike, H. Information theory and an extension of the maximum likelihood principle. In *Proceedings 2nd International Symposium on Information Theory*; Petrov and Caski, Eds., 1974; pp 267-281
- [66] M. Baraket, *Signal Detection and Estimation*, 2nd ed, Artech House, INC.(2005)
- [67] I. Cohen, Relaxed stastical model for speech enhancement and a priori : SNR estimation , *IEEE, Trans ASSP* 13(5) (2005) 870-881.
- [68] P. Wolf, S, Godsill, W.J. Ng, Bayesian Variable Selection And Regularization For Time-Frequency Surface Estimation, *J. R. Stat. B66* (2004) 575-589.
- [69] I. Cohen, On The Decision-Direct Estimation Approach Of Ephraim And Malah, In *Proc, IEEE, ICASSP*. (2004) 293-296.
- [70] P. Wolf, S, Godsill, Simple Alternative To The Ephraim And Malah Suppression Rule For Speech Enhancement, *Proc. 11th IEEE. Workshop. Stat. SP* (2001) 201-204.
- [71] T. Lotter, P. Vary, Speech Enhancement By Maximum A Posteriori Spectral Amplitude Estimation Using A Supergaussian Speech Model, *EURASIP, J. Appl. SP* 5(7) (2005) 1110-1126.
- [72] A. Accardi, R. Vox, A Modular Approach To Speech Enhancement With An Application To Speech Coding, In *Proc. IEEE. ICASSP*. (2003) 201-204.
- [73] P. Wolf, S, Godsill, Towards A Perceptually Optimal Spectral Amplitude Estimation For Audio Signal Enhancement, In *Proc. IEEE. ICASSP*. (2000) 821-824.

- [74] R. E. Madsen, L.K.Hansen, O. Winther, Singular Value Decomposition and Principal Component Analysis, (2004)
- [75] J. Rissanen, Modeling By Shortest Data Description. *Automatica* 14 (1978) 465- 471.
- [76] J. Shlens, A Tutorial on Principal Component Analysis, Systems Neurobiology Laboratory, Salk Institute for Biological Studies La Jolla, CA 92037 and Institute for Non Linear Science, University of California (2005) CA 92 093-0402.
- [77] D. W. Tufts, R. Kumaresan, I. Kirsteins, Data Adaptive Signal Estimation By Singular Value Decomposition Of A Data Matrix, In Proc. IEEE. 70 (6) (1982) 684-685.
- [78] J. A. Gadzow, Signal Enhancement- A Composite Property Mapping Algorithm, IEEE. Trans. ASSP. 36 (1) (1998) 49-62.
- [79] S. Van. Huffel, Enhancement Resolution Based On Minimum Variance Estimation And Exponential Data Modeling, SP 33(3) (1993) 333-355.
- [80] A. Rezaee, Gazor, An Adaptive KLT Approach For Speech Enhancement, IEEE Trans. SAP 9 (2) (2001) 87-95.
- [81] J. Shlens, A Tutorial on Principal Component Analysis, Systems Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla CA 92037 and Institute for Nonlinear Science, University of California, San Diego La Jolla, (2005) CA92093-0402
- [82] J. Shlens, A Tutorial On Principal Component Analysis Derivation, Discussion And Singular Value Decomposition, http://www.cs.princeton.edu/picasso/mats/PCA-Tutorial-Intuition_jp.pdf (2003).
- [83] Zwick, W. R.; Velicer, W. F. Comparison Of Five Rules For Determining The Number Of Components To Retain. *Psychol. Bull.* 99 (3) (1986) 432-442.
- [84] S.J. Qin, R. Dunia, Determining The Number Of Principal Components For Best Reconstruction, *Journal Of Process Control.* 10 (2000) 245–250.
- [85] S. Valle, W. Li, S.J. Qin, Selection Of The Number Of Principal Components: The Variance Of The Reconstruction Error Criterion With A Comparison To Other Methods, *Industrial and Engineering Chemistry Research* 38 (1999). 4389-4401.
- [86] M. Wax, T. Kailath, Detection of signals by information criteria. *IEEE Trans. ASSP.* 33 (1985) 387-392.
- [87] E. R. Malinowski, Determination Of The Number Of Factors And The Experimental Error In A Data Matrix. *Anal. Chem.* 49 (4) (1977) 612-617.
- [88] IEEE Subcommittee (1969), IEEE Recommended Practice For Speech Quality Measurements, *IEEE Trans, Audio Electroacoust.* AU-17 (3) 225-246
- [89] ITU-T (2003), Subjective Test Methodology For Evaluating Speech Communication Systems That Include Noise Suppression Algorithm, ITU-T Recommendation P.835.
- [90] D. Richards, Speech Transmission Performance Of PCM Systems, *Electron Letters.* 1 (1965) 40-91.
- [91] ITU-T (2000), Perceptual Evaluation Of Speech Quality (PESQ) And Objective Method For End-To-End Speech Quality Assessment Of Narrowband Telephone Networks And Speech Codecs, Recommendation P.862.
- [92] D. Klatt, Prediction Of Perceived Phonetic Distance From Critical Band Spectra, In Proc. IEEE. ICASSP 7 (1988) 1278-1281.
- [93] S. Quackenbush, T. Barnwell, M. Clements, Objective Measure Of Speech Quality, Englewood Clift, NJ : Prentice Hall
- [94] A. Abolhassani, S.A. Selouani, D. O'Shaughnessy, Subspace-Based Speech Enhancement by Updating Noise Characteristics in the Presence of Speech, 16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP.
- [95] Y. Benahmed, S.A. Selouani, D. O'Shaughnessy, A. Abolhassani, Real-Life Speech-Enabled System To Enhance Interaction With RFID Networks In Noisy Environments, In proc.IEEE, ICASSP (2011) 1781- 1784
- [96] S.A. Selouani, Speech Processing and Soft Computing, Springer Briefs In Electrical and Computer Engineering, (2011).

- [97] Y. Hu, P.C. Loizou, Evaluation of Objective Quality Measures for Speech Enhancement, *IEEE Trans. on ASLP*. 16(1) (2008) 229–238.
- [98] A. Saadoune, A. Amrouche, M. Debyeche, Rehaussement de la Parole par les Méthodes Spectrales et Statistique, *Traitement et Analyse de l'Information Méthodes et Applications*, TAIMA'2011, (Tunisie) 2011.
- [99] A. Saadoune, A. Amrouche, S.A. Selouani, Speech enhancement in the KLT Domain using MCRA Noise Estimation and the Variance of the Reconstruction Error, 8^{ème} Conférence sur le Génie Electrique, CGE'08, (2013) Algérie.
- [100] J. Dompierre, Properties of Singular Value Decomposition, Department of Mathematics and Computer Science Laurentian University (2009).
- [101] Paliwal, A speech enhancement method based on Kalman filtering, *ICASSP 12* (1987) 177-180.
- [102] J. D. Gibson, B. Koo, and S. D. Gray, Filtering of colored noise for speech enhancement and coding, *IEEE Trans. Signal Processing*, 39 (1991) 1732–1742.
- [103] Weiziu. Du, Driessen, P, Speech enhancement based on Kalman filtering and EM algorithm, *Communications, Computers and Signal Processing*, IEEE 1 (1991) 142 - 145